July 2018

Skybridge-3D-CMOS: A Fine-Grained Vertical 3D-CMOS Technology Paving New Direction for 3D IC

Jiajun Shi
Electircal and Computer Engineering

SKYBRIDGE-3D-CMOS: A FINE-GRAINED VERTICAL 3D-CMOS TECHNOLOGY PAVING NEW DIRECTION FOR 3D IC

A Dissertation Presented

by

JIAJUN SHI

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2018

Electrical and Computer Engineering

SKYBRIDGE-3D-CMOS: A FINE-GRAINED VERTICAL 3D-CMOS TECHNOLOGY PAVING NEW DIRECTION FOR 3D IC

A Dissertation Presented

by

JIAJUN SHI

Approved as to style and content by:

_____
Csaba Andras Moritz, Chair

_____
Daniel Holcomb, Member

_____
Zlatan Aksamija, Member

_____
Lorena Anghel, Member

_____
Christopher V. Hollot, Department Head
Electrical and Computer Engineering

ACKNOWLEDGEMENTS

I would like to gratefully and sincerely thank Dr. Csaba Andras Moritz for his guidance, understanding and patience throughout my Phd and Master study at Umass Amherst. His mentorship was paramount in providing me correct attitude in doing research and presentations. For everything you've done for me, Dr. Moritz, I thank you. I would also like to thank the members of my research group, especially Mostafizur Rahman, Santosh Khasanvis, Sachin Bhat, Sourabh Kulkarni and my good partner Mingyu Li for their consistent help during my Phd study. I would like to thank the Electrical and Computer Engineering Department of Umass Amherst for the support, which turns my dream into reality. In addition, I would also thank all members in my Phd Thesis committee, Drs. Csaba Andras Moritz, Daniel Holcomb, Zlatan Aksamija and Lorena Anghel, for their significant help from proposal to the final defense.

I want to thank my parents, Jidong Shi and Aisu Xu, for their selfless and endless love to me from my childhood until now. Thanks for their unconditional support and understanding of my career choices in U.S. I also need to thank my girlfriend Xun Zhang who always encourage me when I was in tough situations and patiently wait me until my accomplishment of Phd study. Also for my uncle Jiqiang Shi and aunt Xiaoying Shi, thanks for his selfless help to my study in U.S and my parents in China. At last, I need to say thanks to all of my friends who helped and guided me in my tough periods.

Thank you, to all of you!

ABSTRACT

SKYBRIDGE-3D-CMOS: A FINE-GRAINED VERTICAL 3D-CMOS TECHNOLOGY PAVING NEW DIRECTION FOR 3D IC

MAY 2018

JIAJUN SHI

B.Eng., UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA, CHENG DU, CHINA

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS, AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS, AMHERST

Directed by: Professor Csaba Andras Moritz

2D CMOS integrated circuit (IC) technology scaling faces severe challenges that result from device scaling limitations, interconnect bottleneck that dominates power and performance, etc. 3D ICs with die-die and layer-layer stacking using Through Silicon Vias (TSVs) [1] and Monolithic Inter-layer Vias (MIVs) [2][3] have been explored in recent years to generate circuits with considerable interconnect saving for continuing technology scaling. However, these 3D IC technologies still rely on conventional 2D CMOS's device, circuit and interconnect mindset showing only incremental benefits [12] while adding new challenges reliability issues [11][16], robustness of power delivery network design [40] and short-channel effects as technology node scaling [44].

Skybridge-3D-CMOS (S3DC) [18] is a fine-grained 3D IC fabric that uses vertically-stacked gates and 3D interconnections composed on vertical nanowires to yield orders of magnitude benefits over 2D ICs. This 3D fabric fully uses the vertical dimension instead of relying on a multi-layered 2D mindset. Its core fabric aspects

including device, circuit-style, interconnect and heat-extraction components are co-architected considering the major challenges in 3D IC technology. In S3DC, the 3D interconnections provide greater routing capacity in both vertical and horizontal directions compared to conventional 3D ICs [8][23][24][38], which eliminates the routability issue in conventional 3D IC technology while enabling ultra-high density design and significant benefits over 2D. Also, the improved vertical routing capacity in S3DC is beneficial for achieving robust and high-density power delivery network (PDN) design while conventional 3D IC has design issues in PDN design due to limited routing resource in vertical direction. Additionally, the 3D gate-all-around transistor incorporating with 3D interconnect in S3DC enables significant SRAM design benefits and good tolerance of process variation compared to conventional 3D IC technology as well as 2D CMOS.

The transistor-level (TR-L) monolithic 3D IC (M3D) is the state-of-the-art monolithic 3D technology which shows better benefits than other M3D approaches as well as the TSV-based 3D IC approach. The S3DC is evaluated in large-scale benchmark circuits with comparison to TR-L M3D as well as 2D CMOS. Skybridge yields up to 3x lower power against 2D with no routing congestion in benchmark circuits while TR-L M3D only has up-to 22% power saving with severe routing congestions in the design. The PDN design in S3DC shows <5% IR drop while the PDN design in TR-L M3D has sever IR-drop which is out of standard IR drop budget. The SRAM design in S3DC shows 8x static power efficiency over TR-L M3D and 2D CMOS and significantly improved tolerance in lithography variation.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ..... **IV**

ABSTRACT ..... **V**

LIST OF TABLES ..... **IX**

LIST OF FIGURES ..... **X**

CHAPTER **1**

1. INTRODUCTION AND MOTIVATION **1**

1.1 3D IC Technology and Key Issues 2
1.2 Skybridge 3D Fabric 6

2. OVERVIEW OF SKYBRIDGE-3D-CMOS **8**

2.1 Core Fabric Components and Elementary Circuits 8
2.1.1 Vertical Silicon Nanowires 8
2.1.2 Vertical Gate-All-Around Transistor 10
2.1.3 Ohmic Contacts 12
2.1.4 Coaxial Routing Structures 13
2.1.5 Bridges 15
2.2 Circuit Style and Interconnect 15

3. CAD FLOW FOR DEVICE-TO-SYSTEM CO-DESIGN **18**

3.1 Device Simulation of VGAA Junctionless Transistors 19
3.2 Characterization and Abstraction of Standard Cell 20
3.3 Imitation of Cell-to-cell Routing in Large-scale Circuits 21
3.4 Evaluation of Key Metrics 22

4. ROUTABILITY IN S3DC vs. TR-L M3D **24**

4.1 Routability Issue in Conventional 3D IC 24
4.2 Routability in S3DC 25
4.2.1 Theoretical Calculation using Rent's rule 26
4.2.2 CAD-based Simulation 29
4.2.3 Full-chip benchmarking (Logic+ Memory) 30
4.3 Evaluation of Key Metrics 31

5. POWER DELIVERY NETWORK DESIGN 33

5.1 PDN Design Issue in Conventional 3D IC 33
5.2 Robust PDN Design in S3DC 34
5.2.1 PDN Design and Major Issue in TR-L M3D 34
5.2.2 PDN Design in S3DC 37
5.2.3 Methodology of PDN Extraction and IR-Drop Evaluation 39
5.2.4 IR-drop Distribution in S3DC vs. TR-L M3D 41
5.3 PDN's Impact on Routing Congestion 43
5.3.1 PDN's Impact on Routing Congestion 43
5.3.2 PDN's impact on Signal Integrity 45

6. SRAM DESIGN AND VARIATION TOLERANCE 50

6.1 Design and Scaling Issues in Conventional 3D SRAM 50
6.2 Design and Benefits in S3DC SRAM 52
6.2.1 SRAM Design in M3D 52
6.2.2 SRAM Design in S3DC 53
6.2.3 Evaluation of Key Metrics 54
6.3 Variation Tolerance in S3DC SRAM 57
6.3.1 Analytical Model of VGAA Transistor 59
6.3.2 Device-to-circuit Simulation for SRAM 63
6.3.3 Variation's Impact on Noise Margin and Failure Rate 65

BIBLIOGRAPHY 69

LIST OF TABLES

Tables **Page**

4.1: Results of Benchmarking……..……32

5.1: Average IR-drop (Unit: mv)……...……...……43

LIST OF FIGURES

Figures **Page**

1.1 I_{off} versus L_{eff} at VDD=1V for bulk-Si and Double-Gate devices implemented inverters……..1

1.2 Overview of TSV-based 3D IC……...…….2

1.3 Structure of G-L, TR-L and B-L M3D……...……..3

1.4 Implementation of DES core in TR-L M3D and 2D CMOS……..……..…4

1.5 overview of S3DC structure…….6

2.1 A) Dual-doped silicon substrate; B) Dual-doped silicon nanowire array…….....9

2.2 A) N-type VGAA junctionless transistor and Ohmic Contact on n-type nanowire connecting with bridge; B) p-type VGAA junctionless transistor and Ohmic Contact on p-type nanowire connecting with bridge; C) Coaxial routing structure with inter-region contact region; D) Experimental demonstration of vertical Si nanowire array (500nm Height); E) Experimental demonstration of vertical Si nanowire with 400nm height and 20nm width…….10

2.3 A) Drain current vs. drain voltage (I_{DS} - V_{DS}) curve of n-type device; B) Drain current vs. drain voltage (I_{DS} - V_{DS}) curve of p-type device…….……...11

2.4 Gate capacitance vs. gate voltage (C_g – V_g) curve of n-type device with V_{DS}=0.2V-0.8V and V_S=0V.....12

2.5 A) Fabricated metal (Ni) to silicon (n-type) Ohmic contact; B) Fabricated bridge on planarized inter-layer dielectric……...……13

2.6 Simulated I-V curve of inter-region contact structure……14

2.7 A) structure of coaxial routing; B) 3D layout of NAND3 gate in S3DC; C) Interconnections between vertical 3D gates in S3DC……..15

3.1 Skybridge-3D IC device-to-system design flow……...……18

3.2 TCAD device simulation: A) Generated n-type VGAA structure with high-density meshing [22] in channel, gate oxide and gate metal regions; B) Uniform heavy doping (10^{20} cm^{-3}) in S/D and channel for our n-type VGAA transistor……...…...19

3.3 A) Layout of NAND3 cell in S3DC; B) Abstracted LEF format of S3DC NAND3 cell……..……..20

3.4 A) Schematic of a sample circuit with three NAND2 gates and one NAND3 gate; B) Placement of the sample circuit; C) Layout of the implementation of the sample circuit based on S3DC…….…….21

4.1 A) Normalized routing demand in 2D CMOS, TR-L M3D, S3DC; B) Routing demand/resource ratio in all technologies' LDPCs……..28

4.2 Layouts of LDPC in 2D CMOS, TR-LM3D and S3DC……...…….29

4.3 Layouts of LDPC in 2D CMOS, TR-LM3D and S3DC……...…….31

5.1 A) Low-density PDN design in the typical TR-L M3D; B) High-density PDN design in the improved version of TR-L M3D…….35

5.2 A) PDN design in S3DC; B) S3DC's PDN routing implemented in Cadence Encounter…….38

5.3 Current density distribution in Sentaurus TCAD simulation of silicided vertical routing nanowire …….41

5.4 IR-drop distribution in AES benchmark simulated in Cadence Voltus: A) Top-tier of TR-L M3D with high density PDN; B) Bot-tier of TR-L M3D with high density PDN; C) S3DC……42

5.5 Routing congestion comparison of AES benchmark of TR-L M3D with and without PDN (low-density PND version)……44

5.6 A) M2 Density in 5µm* 5µm Square (TR-L M3D); B) M2 Density in 5µm* 5µm Square (2D CMOS)…….…….46

5.7 Methodology of SI analysis……47

5.8 A) SI Results in 2D CMOS; B) SI Results in TR-L M3D; C) SI Results in S3DC……48

5.9 PDN's help in SI improvement in S3DC……...……49

6.1 A) Top-tier (PMOS) and bot-tier (NMOS) layout design in M3D-based SRAM [42]; B) Schematic views of a lateral stacked nanowire transistor (left) and a vertical nanowire transistor (right)…….……...51

6.2 Layout of SRAM design in S3DC……..……...……..53

6.3 Read and Write NM of each technology based 6T-SRAM……...…….54

6.4 Read and Write time of each technology based 6T-SRAM……56

6.5 A) Leakage of each technology based 6T-SRAM; B) Comparison for TCAD based model vs. our analytical model in read NM's evaluation……..……...57

6.6 Contour lines of electrostatic potentials inside channel……...……..60

6.7 Quasi-Fermi potential distribution at the center of the channel……...……..61

6.8 A) Modeled I-V vs. TCAD simulated for various channel widths; B) Modeled I-V vs. TCAD simulated for various channel lengths……...……..62

6.9 Evaluation flow of variation impact on SRAM……..63

6.10 Gaussian distribution of read NM for channel length variation……...…65

6.11 Gaussian distribution of read NM for channel width variation……...…….65

6.12 A) (μ-6σ) criterion for channel length variation in S3DC's SRAM as channel width varies from -3σ to +3σ; B) (μ-6σ) criterion for channel length variation in M3D and 2D CMOS based SRAM as channel width varies from -3σ to +3σ……..66

6.13 A) (μ-6σ) criterion for channel width variation in S3DC's SRAM as channel length varies from -3σ to +3σ; B) (μ-6σ) criterion for channel width variation in M3D and 2D CMOS based SRAM as channel length varies from -3σ to +3σ…….67

CHAPTER 1

INTRODUCTION AND MOTIVATION

Tremendous progress in miniaturization of integrated circuits (ICs) has been crucial for the socio-economic developments in the last century. So far, this miniaturization was mainly enabled by the ability to continuously scale the CMOS technology. As the scale of CMOS technology nodes goes down, it is faced with several challenges and special difficulty to maintain the traditional way of scaling. Firstly, as more transistors integrated into the same die area, it becomes difficult to design compact circuits and routings. Large resistance and capacitance from interconnections cause significant degradation in circuit's performance and power. Microprocessor's performance is faced with a corner and taken into a bottleneck [50]. And the power density of a

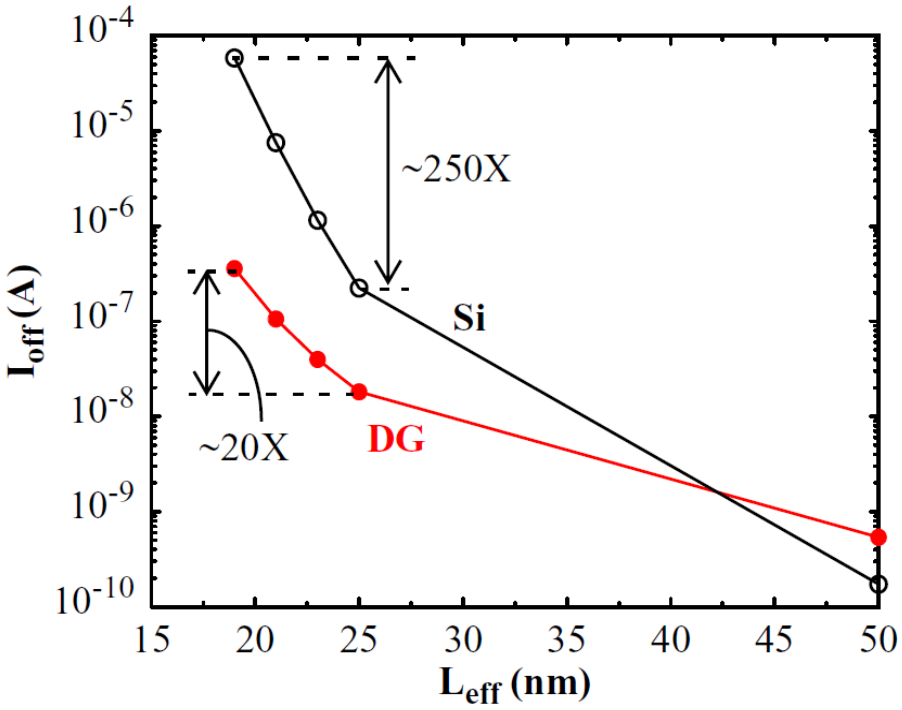

Figure 1.1 **I_{off} versus L_{eff} at VDD=1V for bulk-Si and Double-Gate devices implemented inverters [50].**

microprocessor will soon climb beyond the capabilities of any possible cooling techniques in the future. Secondly, in terms of the devices, technology scaling enhances short channel effects, resulting in the larger off-leakage current. What's more, as the device scales down, the threshold voltage and Vdd value do not go down linearly [50] (See Fig. 1.1), which results in degradations of performance and power in building circuits with high density.

1.1 3D IC Technology and Key Issues

3D IC technology is an alternative pathway for future technology scaling. The main goal of 3D IC technology is to fully use the vertical dimension for compact routing and parasitic reduction over 2D CMOS. With the extensive research on through-silicon-via (TSV) [1] and monolithic 3D ICs (M3D) [2][3] from both academia and industry, mainstream production of 3D ICs is expected in a near future. The conventional 3D IC technology start from TSV-based 3D IC technology in which the logic and memory are integrated in to two separate dies and bonded using conventional packing technology (See Fig. 1.2). The TSV-based 3D IC technology

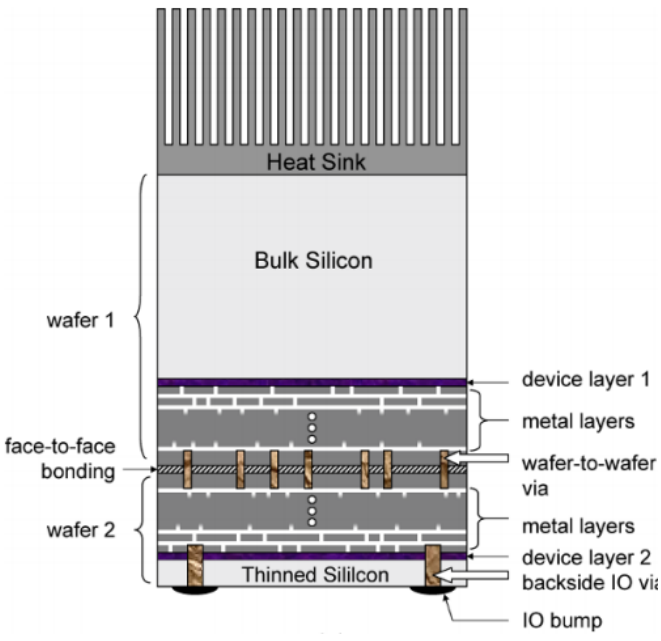

Figure 1.2 **Overview of TSV-based 3D IC [4]**

only implements a coarse-grained 3D interconnection between logic and memory since the via size is relatively large (>10μm) and limited by wafer bonding precision. Therefore, the interconnect saving and followed power and performance benefits in TSV-based 3D IC against 2D CMOS is incremental. Monolithic 3D IC is more advanced 3D IC technology than TSV-based 3D ICs which shows considerable (up-to 20%) power saving and significantly improved (up-to 10%) performance against 2D CMOS. It is an emerging technology which is enabled by sequential vertical integration of extremely thin device layers with very high alignment precision. Unlike TSV-based 3D IC, monolithic inter-layer vias (MIVs) are miniscule (<100nm diameter) and can be used in large numbers within the design. This helps in high integration density allowing numerous 3D connections which results in reduced wirelength, improved power and better performance [11]. The side-view of a typical two-tier monolithic stacking structure with seven metal layers in each tier is shown in Fig. 1.3. The device layer thickness is around 30nm and the inter tier dielectric (ILD) which separates different

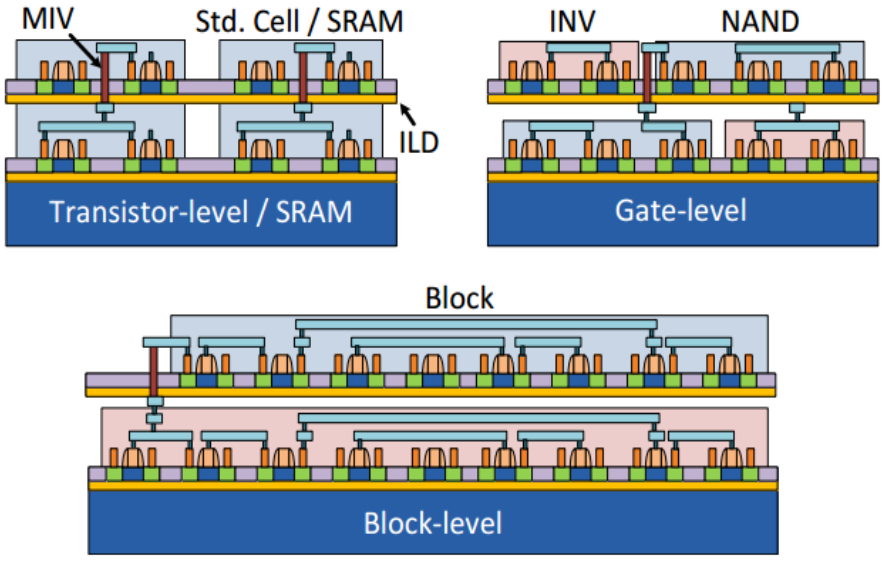

Figure 1.3 **Structure of G-L, TR-L and B-L M3D [13]**

tiers is about 100nm thick.

There are three different design styles in M3D: transistor-level (TR-L) (Fig. 1.3A), gate-level (G-L) (Fig. 1.3B), and block-level (B-L) (Fig. 1.3C) M3D design. TR-L M3D design splits PMOS and NMOS into two tiers within a standard cell, and uses MIVs for intra-cell and inter-cell connections. It is the most fine-grained M3D design style, but takes significant effort because it requires completely new cell GDS layouts containing challenges in the power delivery network (PDN) design. Gate-level M3D design, which is the focus of this paper, utilizes existing cells and places cells into tiers, using MIVs only for inter-cell connections. In block-level M3D design, functional blocks are floorplanned into multiple tiers. However, due to its coarse granularity, there is limit on fine-grained vertical integration.

Among all 3D IC approaches, transistor-level monolithic 3D IC [16] represents the state-of-the-art M3D that uses 3D standard cells for high-density IC design. But it still follows conventional 2D CMOS's routing mindset for inter-cell connections where

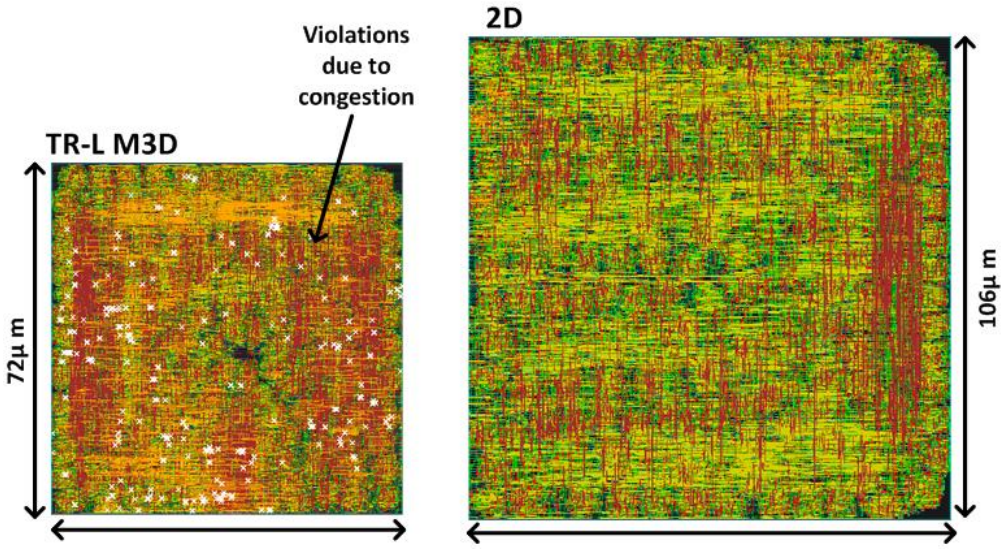

Figure 1.4 **Implementation of DES [36] core in TR-L M3D and 2D CMOS**

the standard cells are placed and routed in a two-dimensional plane limiting their accessibility and routability. This in turn causes severe routing congestion [16] in large-scale TR-L M3D ICs diminishing the benefits of this approach and limiting scalability (See Fig. 1.4).

The design for power-delivery network (PDN) is another major challenge in M3D which is caused by the routability issue. Due to limited routing capacity in vertical direction, PDN on top metal layers has poor accessibility to the device layer away from the power source. This leads to severe IR-drop in this device layer. In gate-level (G-L) M3D IC [11][40], large number of MIVs need to be used in cell-to-cell communication between top- and bot-tier while limited number of MIVs are used in the PDN's vertical routing to the bot-tier. Therefore, taking some cell-to-cell routing resources for PDN routing or enlarging design area to add routing resource for PDN, is the only way to achieve a robust and high-density PDN design in G-L M3D [40]. In the typical version of transistor-level (TR-L) M3D [4][16], top-tier's high-density routing creates blockages, which limit PDN's vertical routing access to bot-tier and results in an incomplete and low-density PDN design. In the improved TR-L M3D version [4], larger cell footprint is used to add additional vertical routing resource for PDN's access to bot-tier. Overall, in both G-L and TR-L M3D approaches, the insertion of a robust PDN design would impact 3D cell-to-cell routing density which in turn diminishes the benefits over 2D design.

In terms of device, the M3D still uses the conventional tri-gate bulk transistor which has inherent short-channel effects and device reliability issues as 2D CMOS.

Therefore, as technology node scales, the M3D faces the same issue that happens to 2D CMOS which limits M3D's overall benefits compared to other emerging technology directions. Additionally, the short-channel effects result in susceptibility to device geometric variations which usually caused by lithography variation.

1.2 Skybridge 3D Fabric

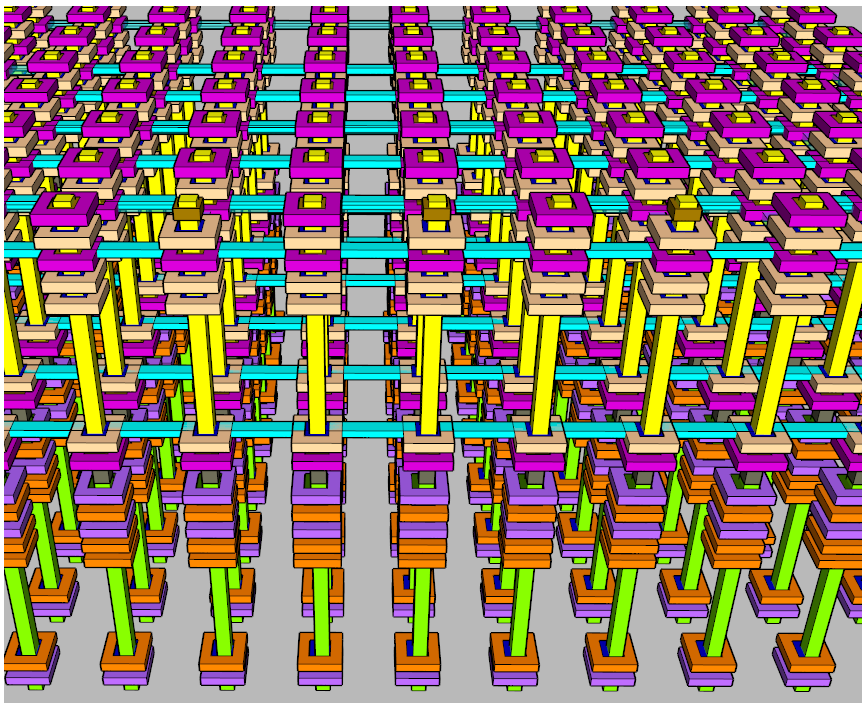

Figure 1.5 **Overview of S3DC structure**

Skybridge 3D CMOS (S3DC) is a recently proposed fine-grained 3D IC fabric relying on vertical nanowires (See Fig. 1.5) that presents a paradigm shift for scaling, while addressing critical challenges in 3D IC technology. Core fabric aspects including device, circuit-style, connectivity [8], thermal management [7] and pathway of manufacturing [9] are co-architected for 3D compatibility. Input/output pins for each vertically-composed gate have multiple points of access both horizontally and

vertically which can be reached through architected routing components, as opposed to TR-L M3D which limits pin-access to a 2D plane and relies on conventional routing schemes. Thus Skybridge fully utilizes the vertical dimension providing increased routability to address high-density routing in large-scale ICs. In S3DC, the greater routing capacity in both vertical and horizontal directions compared to conventional 2D and 3D ICs [8][38], which enables its ultra-high density design and significant benefits over 2D. Also, the improved routing capacity in S3DC is beneficial for a robust and high-density PDN design whose presence would not impact or create blockages on the 3D cell-to-cell routing. Moreover, the use of gate-all-around vertical transistor in S3DC helps in eliminating the short-channel effects in the device. Also, the device channels length control is deposition dependent which has smaller variation than the tri-gate bulk transistor in 2D CMOS. These lead to S3DC's better tolerance of variation compare to the M3D as well as 2D CMOS.

CHAPTER 2

OVERVIEW OF SKYBRIDGE-3D-CMOS

Skybridge-3D-CMOS (S3DC) follows a fabric-centric mindset to create a truly fine-grained 3D integration system in the vertical dimension. Each core component is designed for 3D compatibility and overall system efficiency. These components are assembled on a 3D uniform template of single crystal nanowires that act as scaffolding for vertical assembly. Fig. 1.4 shows the envisioned S3DC; Using a similar process flow described in [5], vertical nanowires, are constructed primarily through masking and high aspect ratio etching on heavily doped silicon bulk (other methods are also possible). Architected fabric components are constructed on these nanowires by using material deposition techniques [5]. In this section, we present the core components that enable fine-grained integration of both n- and p-type nanowires in S3DC. Detailed explanation of material selection and working mechanism are presented to illustrate how these components are used in unison to achieve desired functionality and 3D compatibility with circuits implemented across both horizontal and vertical dimensions.

2.1 Core Fabric Components and Elementary Circuits

2.1.1 Vertical Silicon Nanowires

Vertical nanowires are the fundamental building blocks that enable vertical stacking of designed core Skybridge components. The nanowires serve multiple functions – they can act as (i) logic nanowires that have stacked transistors to implement required

logic gates, (ii) routing nanowires to carry electrical signals along the vertical dimension, and (iii) heat dissipating nanowires to extract and sink heat generated during circuit operation to the bulk substrate [7].

The nanowire formation step precedes all manufacturing steps, and is done after wafer preparation. Wafer preparation involves stacking heavily doped n-type and p-type silicon layers to create a dual-doped silicon wafer (Fig. 2.1A). This can be achieved by bonding heavily doped n-type and p-type substrates using techniques that are similar to the ones described in literature [2][3] and currently used for conventional 3D ICs. A silicon dioxide layer is used between the n-type and p-type doped silicon layers for isolation. Vertical nanowire patterning can be achieved through inductively coupled plasma etching (~50:1 aspect ratio, 5nm dimension shown) [5] and has been experimentally demonstrated as shown in Fig. 2.2D-E.

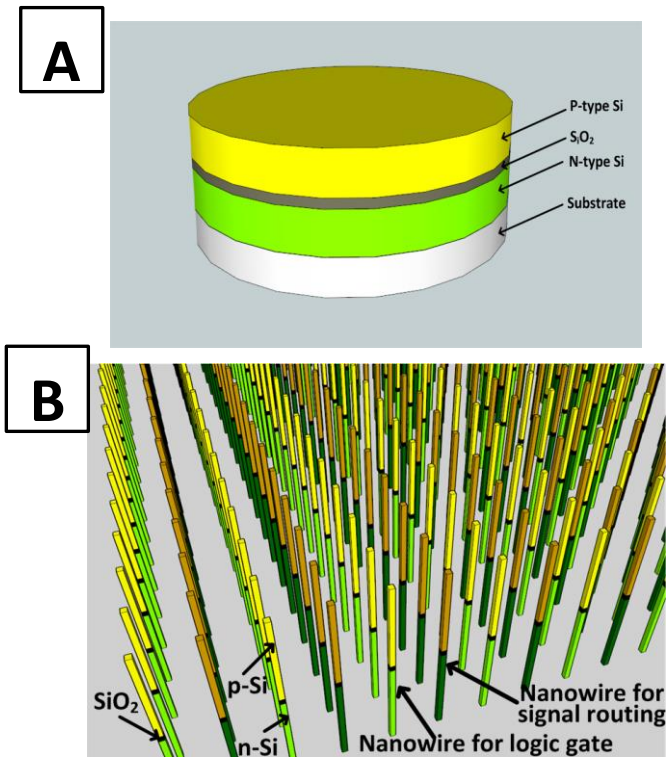

Figure 2.1 **A) Dual-doped silicon substrate; B) Dual-doped silicon nanowire**

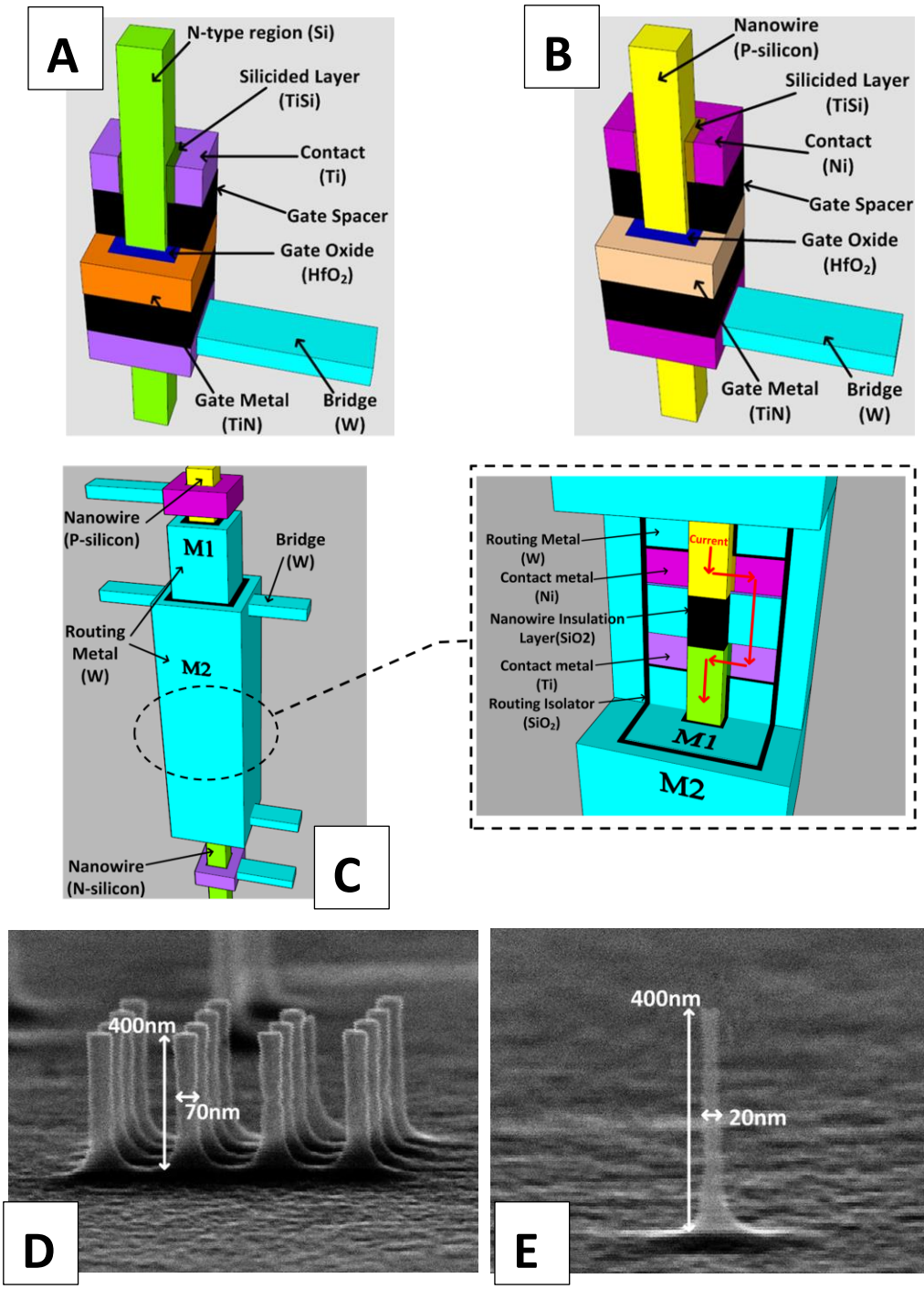

Figure 2.2 **A) N-type VGAA junctionless transistor and Ohmic Contact on n-type nanowire connecting with bridge; B) p-type VGAA junctionless transistor and Ohmic Contact on p-type nanowire connecting with bridge; C) Coaxial routing structure with inter-region contact region; D) Experimental demonstration of vertical Si nanowire array (500nm Height); E) Experimental demonstration of vertical Si nanowire with 400nm height and 20nm width**

2.1.2 Vertical Gate-All-Around Transistor

VGAA junctionless transistors are used as active devices, and are formed on

nanowires through consecutive material deposition steps [51]. These junctionless transistors use uniform doping with no abrupt variation in Drain/Source/Channel regions (Fig 2.2A-B), which simplifies manufacturing requirements and is especially suitable for this fabric. Their channel conduction is modulated by the workfunction difference between the heavily doped channel and the gate [41]. Titanium Nitride

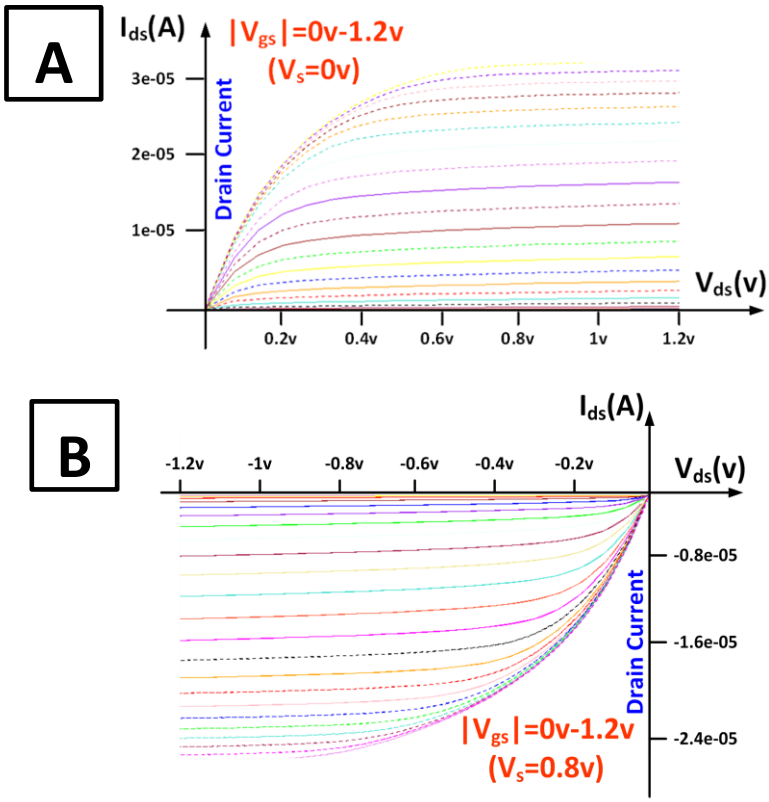

Figure 2.3 **A) Drain current vs. drain voltage (I_{DS} - V_{DS}) curve of n-type device; B) Drain current vs. drain voltage (I_{DS} - V_{DS}) curve of p-type device**

(TiN) and Tungsten Nitride (WN) are chosen for n-type and p-type transistors respectively to provide the required workfunction for the accumulation mode when the transistor is ON [26][27]. 3D TCAD Process and Device simulations [20] were used to extract the device I-V characteristics, shown in 2.3A. The n-type device had an ON current of 30 µA, and OFF current 0.1nA. The p-type device had an ON current of 26 µA, OFF current 0.76nA. 2.4 shows the TCAD-simulated gate capacitance of the

n-type VGAA transistor with applying various Vds values. In saturation state, the VGAA transistor has around 250aF gate capacitance. The simulation methodology and assumptions are detailed in Chapter 3.

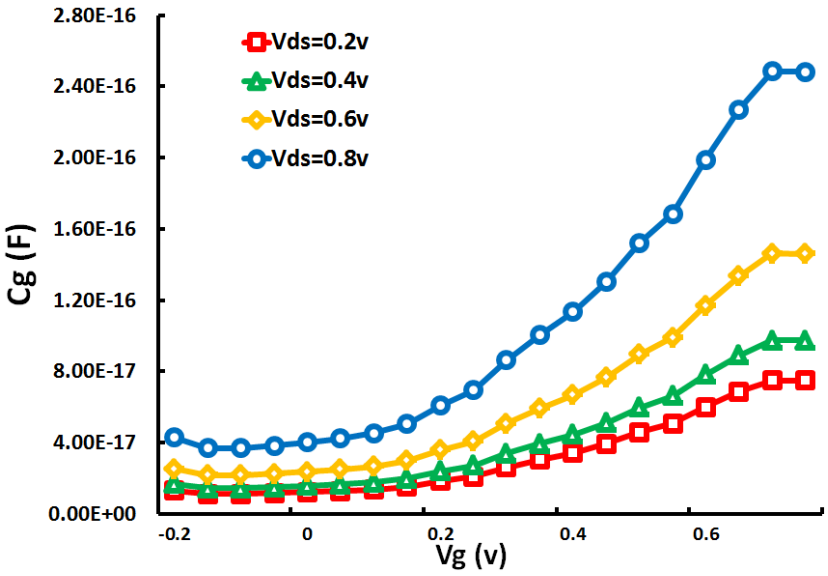

Figure 2.4 **Gate capacitance vs. gate voltage (C_g – V_g) curve of n-type device with V_{DS}=0.2V-0.8V and Vs=0V**

2.1.3 Ohmic Contacts

In S3DC the input/output ports of different gates are connected using horizontal metallic routing components called bridges (See Chapter 2.1.5) and vertical coaxial routing structures (See Chapter 2.1.4). Specific materials are chosen for each doped silicon region to minimize contact resistance between heavily-doped silicon and metals (Fig. 2.2A-B). Nickel is used for creating a low-resistance Ohmic contact with p-doped silicon and Titanium is chosen for n-doped silicon. Each of these metals has the proper workfunction to eliminate Schottky Barrier in the interface with corresponding doped silicon, achieving low resistance; in addition, they also have good adhesion to doped silicon [26][27]. A thin Titanium Nitride layer in the p-type nanowire Ohmic contact is used for avoiding the chemical reaction between Nickel

and Tungsten. 2.5A shows experimental demonstration of Ohmic contact formation through material deposition around vertical nanowire.

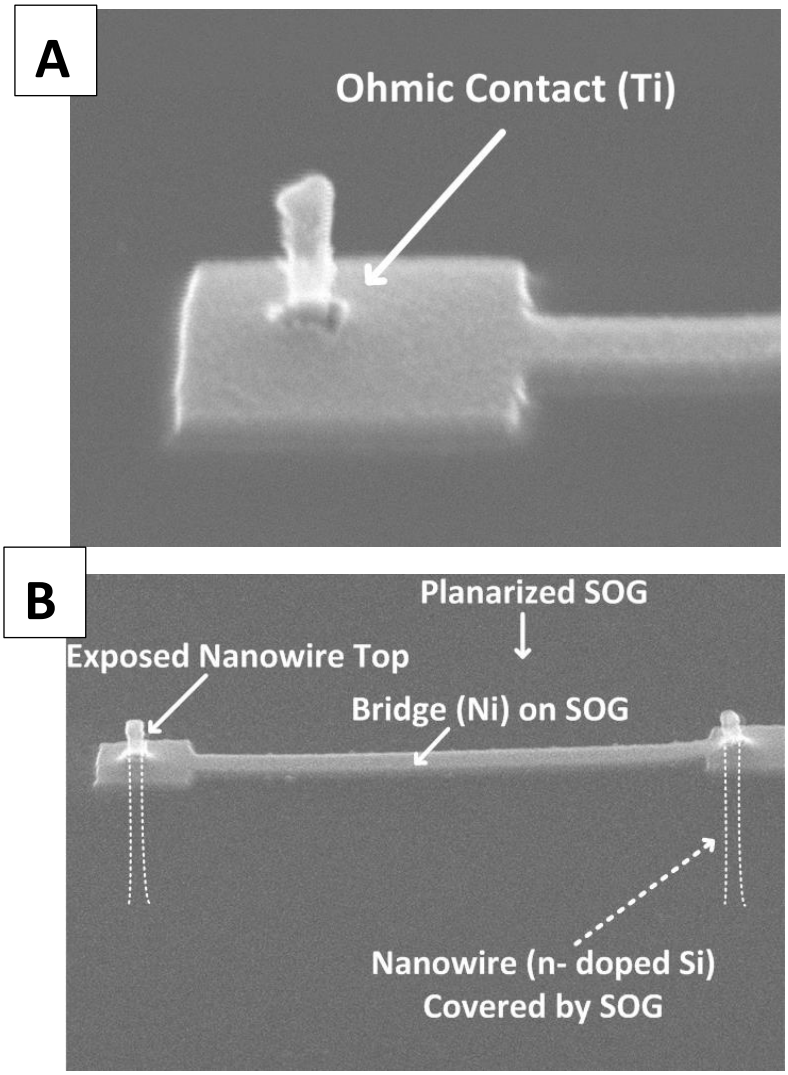

Figure 2.5 **A) Fabricated metal (Ni) to silicon (n-type) Ohmic contact; B) Fabricated bridge on planarized inter-layer dielectric**

2.1.4 Coaxial Routing Structures

Coaxial routing refers to a scheme where an outer signal routing layer runs coaxially with another inner signal routing layer without affecting each other. Every routing layer in such a coaxial structure facilitates signal propagation along the vertical dimension. This is unique and enabled by the fabric's vertical integration approach, and can be manufactured similar to the process flow used in ref. [51]. A

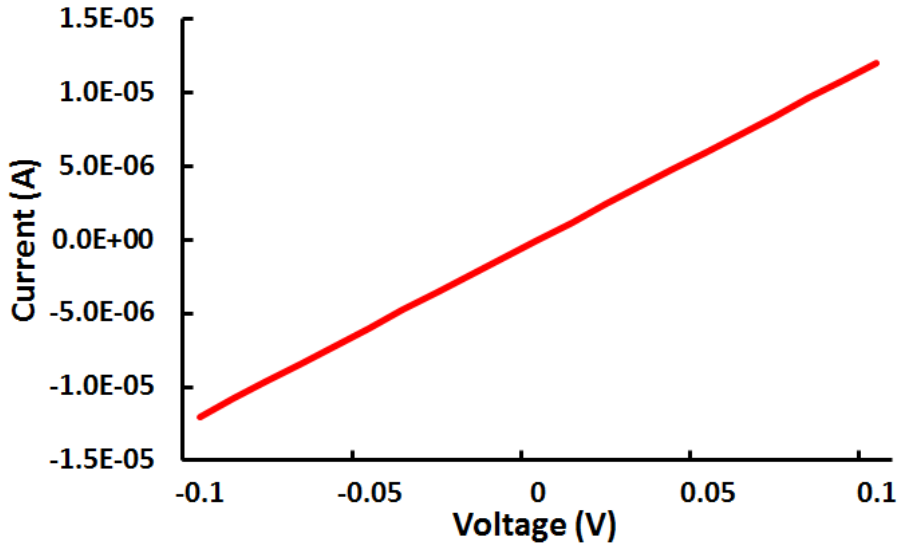

Figure 2.6 **Simulated I-V curve of inter-region contact structure**

coaxial routing structure (Fig. 2.2C) consists of two concentric metal layers separated by dielectric layers around a nanowire. The outermost metal shell (M2) and the inner nanowire are used for carrying input/output signals. Electrical coupling noise between the inner nanowire and outer metal shell can be mitigated by pinning the inner metal shell (M1) to a ground (GND) signal for shielding. Fig. 2.2C illustrates this concept; the GND signal is applied to the M1 metal shell which thus acts as a shield layer, and prevents coupling between signals in M2 shell and the inner nanowire.

Given that a nanowire itself can carry a signal and the fabric incorporates both n- and p-type nanowires, it needs support to allow signal routing between n- and p-regions bypassing the isolation dielectric layer between them. An inter-region contact structure is designed for this purpose to form a low resistance Ohmic contact between p-type and n-type regions on a single nanowire (Fig. 2.2C). 2.6 shows the I-V characteristics of the contact structure that was carried out by emulating the fabrication process flow in Synopsys Sentaurus TCAD [20] (see Chapter 3).

2.1.5 Bridges

Bridges (Fig. 2.2A-B) connect with Ohmic contacts and coaxial routing structures to carry and propagate signals horizontally in-between nanowires. As shown in Fig. 2.2A and Fig. 2.2B, Tungsten is used as the material to form the bridges because of its good adhesion with Titanium.

2.2 Circuit Style and Interconnect

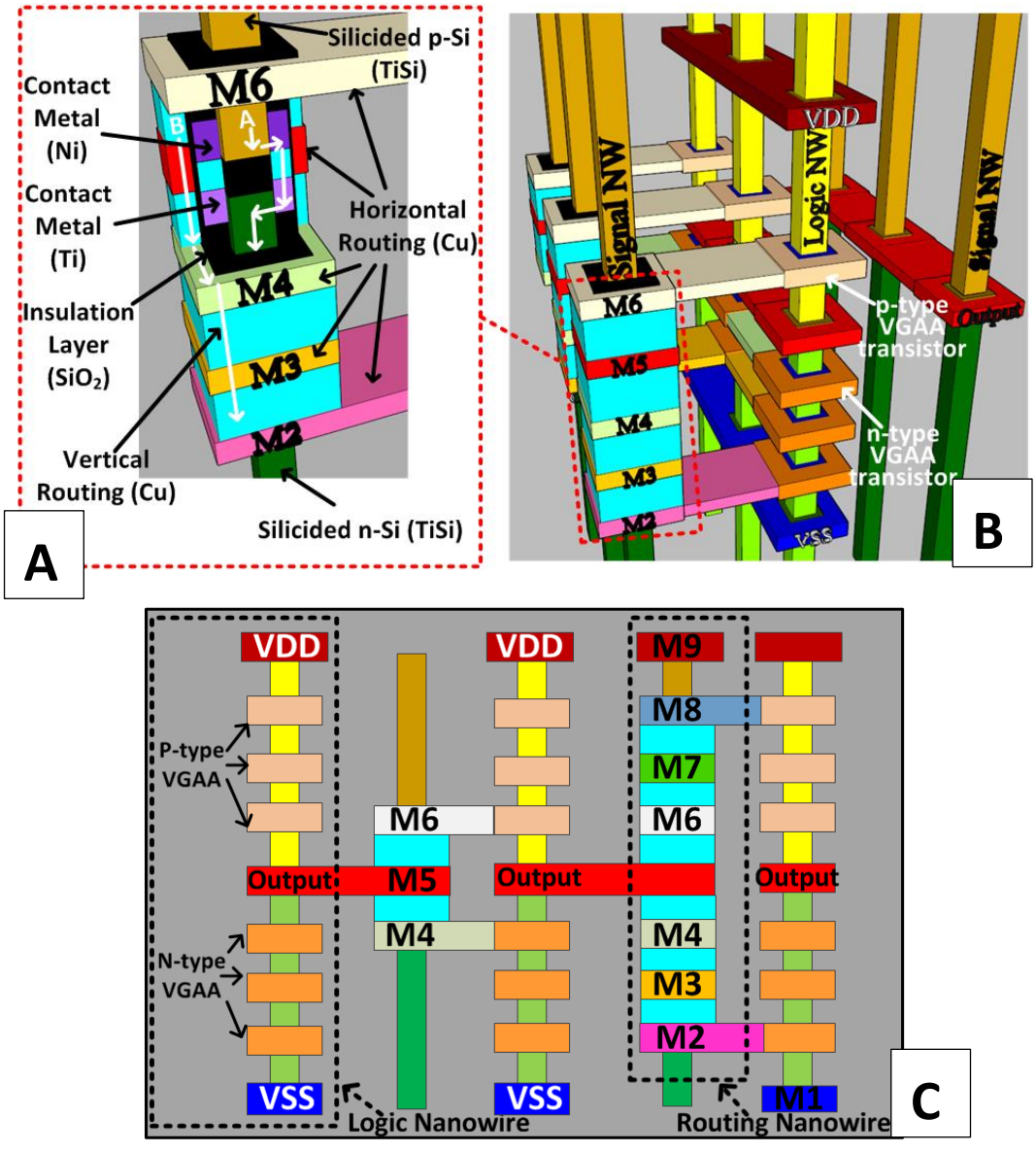

Figure 2.7 **A) structure of coaxial routing; B) 3D layout of NAND3 gate in S3DC; C) Interconnections between vertical 3D gates in S3DC**

As mentioned in last sections, each dual-doped nanowire has p-type doped silicon on the top-half, n-type doped silicon on the bot-half and a dielectric layer in-between for insulation (See Fig. 2.1B). In S3DC fabric design, the nanowire array consists of rows of logic nanowires and rows of routing nanowires (See Fig. 2.1B). The logic nanowire is used in logic gate implementation. Core components including n-type and p-type Vertical Gate-All-Around (n-VGAA and p-VGAA) junctionless transistors [41], are stacked on n-type doped and p-type doped regions of each logic nanowire to implement complementary logics of static-logic gates. In order to create low-resistivity PDN network (See Chapter 5), the routing nanowire has silicided n- and p-type silicon regions (TiSi) for low- resistivity routing. The S3DC fabric is designed with various horizontal metal layers that are vertically stacked along nanowires with uniform thickness and vertical spacing (See Fig. 2.7C).

2.7B shows the layout of a 3-input 3D NAND gate that is built with 9 nanowires. 3 logic nanowires with 6 stacked VGAA transistors are used for logic implementation. 6 routing nanowires with coaxial routing structures are used for creating input/output pins of the NAND3 gate. In total, 9 horizontal metal layers (M1-M9) are used in the design of S3DC standard cell (See Fig. 2.7C): M9 is used to place VDD rails which consist of bridges and bridge-to-nanowire contacts, VSS rails with similar structure are placed in M1, output port is created by M5 with an inner connection to the inter-layer contact structure of logic nanowire, n-VGAA transistors are placed in three layers M2-M5 and p-VGAA transistors are placed in three layers M6-M8. The feature sizes of contact metal, bridge, VGAA transistors and the nanowire pitch are designed

following the design rules as described in [5]. Additional metal layers (M10-M11) are added on the top of nanowires array to provide necessary routing resources in PDN and clock tree design.

CHAPTER 3

CAD FLOW FOR DEVICE-TO-SYSTEM CO-DESIGN

S3DC technology follows the static CMOS circuit style as 2D CMOS, but its interconnect, device and cell-layout design are significantly different from 2D CMOS. Consequently, the commercial CAD tools that are used for physical design flow and evaluation in 2D CMOS are not immediately suitable for S3DC. In order to make these 2D CAD tools support S3DC designs, we propose to represent S3DC physical designs in a way that is compatible with the 2D tools – essentially by finding analogous (by function) concepts in 2D physical layouts to the S3DC fabric structures and setting appropriate constraints. Fig. 3.1 shows the proposed device-to-system design flow for RTL-to-layout design in S3DC: it mainly includes Sentaurus TCAD [20] based simulations of n- and p-type VGAA junctionless transistors, characterization of standard cell timing and power (Lib file), characterization of interconnect capacitance and resistance table (.tch file), RTL synthesis, placement and route for layout generation, power and performance evaluation. It is a modified ASIC design flow that is based on 2D CAD tools but severs for S3DC design.

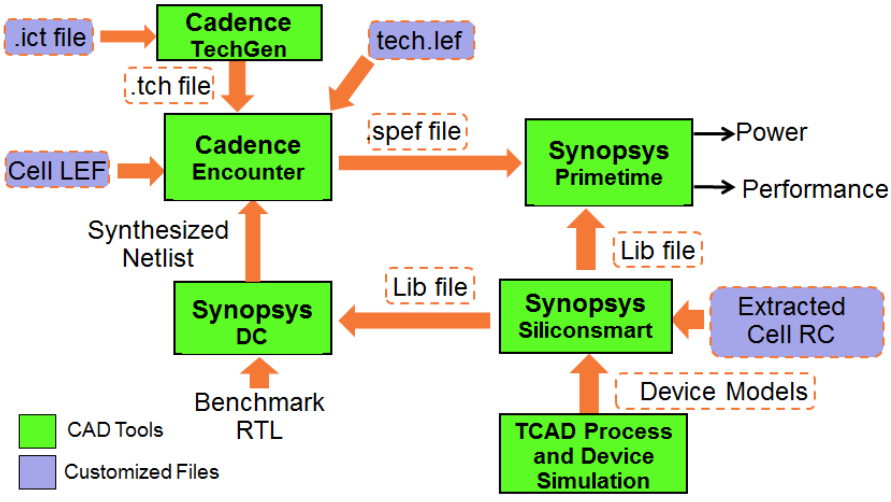

Figure 3.1 **Skybridge-3D IC device-to-system design flow**

3.1 Device Simulation of VGAA Junctionless Transistors

The n-type and p-type VGAA junctionless transistors were extensively characterized using detailed physics-based 3D simulation of the electrostatics and operations using Synopsys Sentaurus TCAD [20]. The Sentaurus Process [20] was used to create the device structure emulating actual process flow; process parameters such as ion implantation dosage, anneal duration and temperature, deposition parameters etc. were similar to our previous experimental process parameters for junctionless device demonstration [9].

The resulting device structure (See Fig. 3.2A) had 16nm long Si channel, 2nm of HfO_2 as gate oxide, 11.5nm thick gate electrode, 5nm long Si_3N_4 as spacer material, and 22nm thick S/D contact material. Gate metal work function is 5.2eV (TiN) and 4.3eV (WN) for n-type and p-type transistors respectively [26][27]. 16nm channel length was simulated following similar feature size as our original Skybridge's device [5]. Uniform doping for drain, channel and source was required to form the VGAA

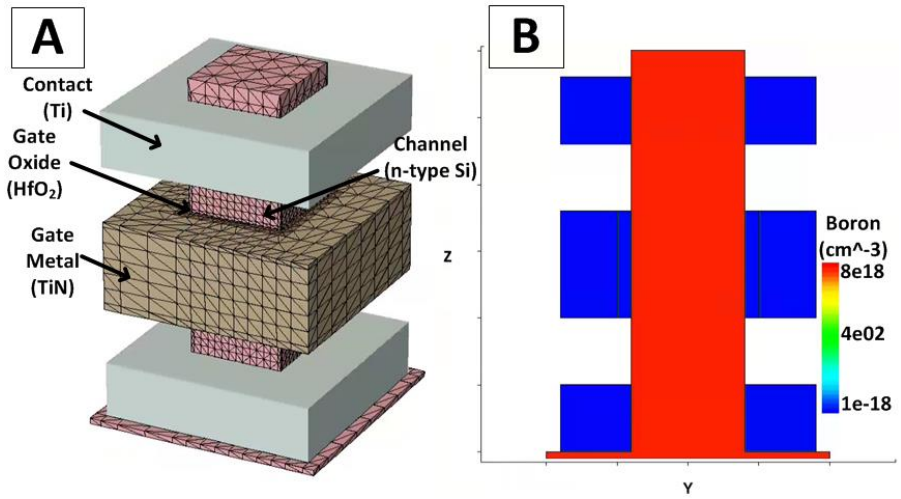

Figure 3.2 **TCAD device simulation: A) Generated n-type VGAA structure with high-density meshing [22] in channel, gate oxide and gate metal regions; B) Uniform heavy doping (10^{20} cm^{-3}) in S/D and channel for our n-type VGAA transistor**

junctionless transistor, and As and Br were chosen as dopants for n- and p-type devices respectively. The doping concentration for n-type device was 10^{19} cm^{-3} and p-type was 10^{20} cm^{-3}. 3D TCAD Device [20] simulations were used to extract the device I-V characteristics. The n-type device had an ON current of 30 µA, and OFF current 0.1nA. The p-type device had an ON current of 26 µA, OFF current 0.76nA.

3.2 Characterization and Abstraction of Standard Cell

We manually designed the standard cell layouts including logic gates, a buffer, and a flip flop, following the S3DC technology design rules [5]. RC extractions of cells were manually done using the Predictive Technology Interconnect Models [31], following the dimensions and material types of the structures in the layouts. Physical HSPICE netlists were then built following the circuit topology and the extracted RC. Synopsys SiliconSmart [22] took the device models and the physical HSPICE netlists as the inputs, and performed power and timing characterization for each standard cell. These results have been written into a cell library file (Lib file) [29], which is used

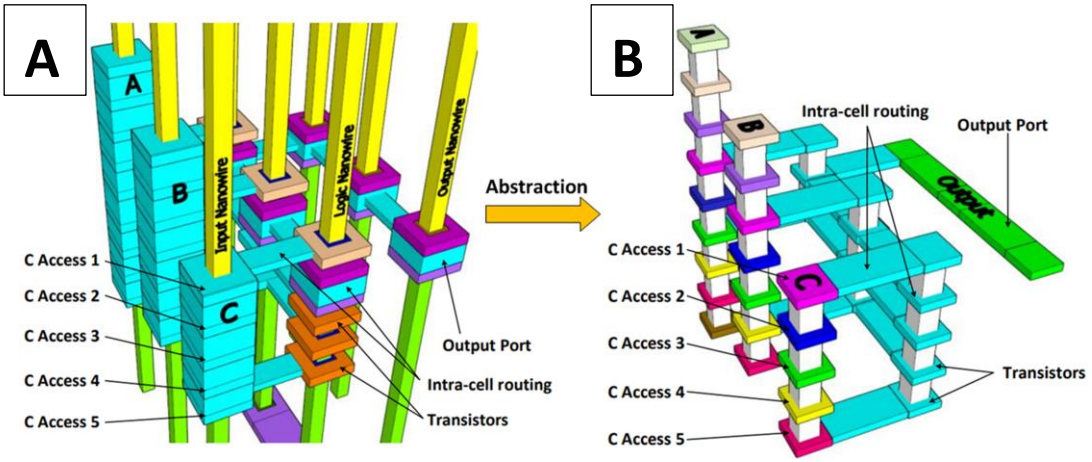

Figure 3.3 **A) Layout of NAND3 cell in S3DC; B) Abstracted LEF format of S3DC NAND3 cell**

during the later design and evaluation stages. The cell Library Exchange Format (LEF) files [28], called cell abstracts, are used in Encounter-based cell-to-cell routing. They contain cell layout information including the dimensions of each cell, the location, layer and dimensions of the pins, and the descriptions of obstructions (the used metal layers / shapes for intra-cell routing). Fig. 3.3A and Fig. 3.3B show the layout design and its LEF abstract of a 3D 3-input NAND gate.

3.3 Imitation of Cell-to-cell Routing in Large-scale Circuits

Cadence Encounter [32] is designed to implement the 2D CMOS layouts. It treats

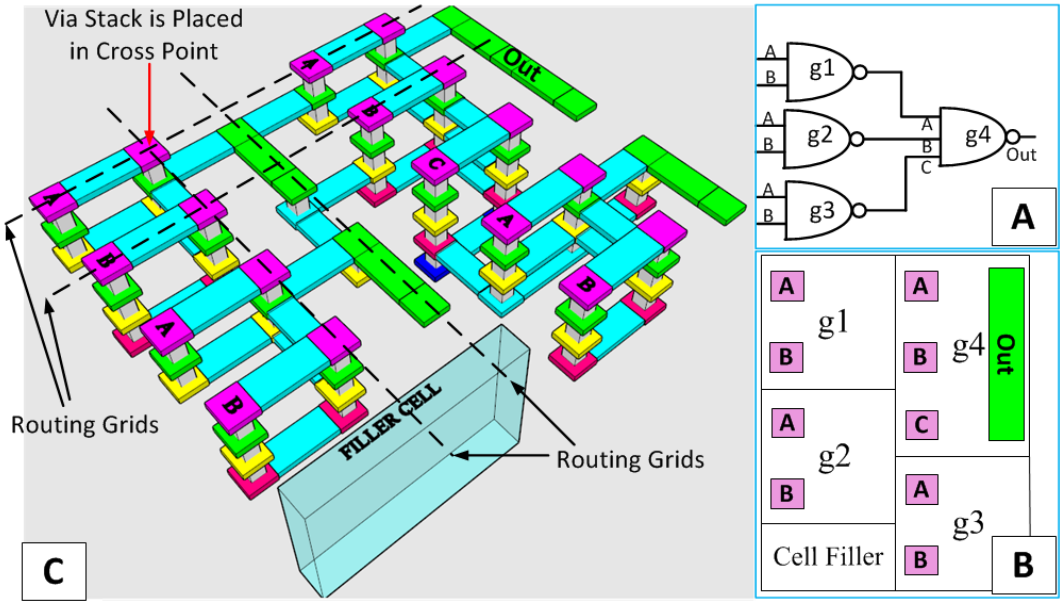

Figure 3.4 **A) Schematic of a sample circuit with three NAND2 gates and one NAND3 gate; B) Placement of the sample circuit; C) Layout of the implementation of the sample circuit based on S3DC**

each standard cell as a black box, only knowing its cell dimensions, and pin and obstruction information from the cell LEF files; it places the cells and routes the nets in such that performance, power, and area are optimized. To make Encounter generate correct S3DC physical designs, in addition to the aforementioned ways to represent

S3DC designs in 2D tools, as is shown in Fig. 3.1, we have added two constraints of inter-cell routing to imitate the S3DC routing style:(i). In S3DC, nanowires are uniformly distributed in an array. The vertical routing, including using Routing Nanowires and Coaxial Routing structures, can only be achieved along these uniformly-distributed nanowires. Consequently, the vias representing these S3DC vertical routing elements in Encounter are only allowed to be placed where the nanowires are positioned in the nanowire array template. (See Fig. 3.4) (ii). The Bridges connect the nanowires and thus are only placed along the tracks defined by the rows/columns of nanowires. So in 2D tools the wires representing these Bridges should only be allowed on the discrete tracks separated by the nanowire pitch in the S3DC template.

All these constraints can be defined in the technology LEF file, which contains the routing rules. Other parameters, including design rules, are also captured in the technology LEF and TCH files. The TCH file [30] sets the inter-cell RC extraction rules, and is generated by Cadence Techgen based on the metal layer design rules. With the cell LEF file, the technology LEF file, and the TCH file, Encounter can imitate the S3DC physical design style, and do the placement and routing for S3DC designs.

3.4 Evaluation of Key Metrics

The key metrics are evaluated by using Synopsys Primetime with imported .spref file, Lib file and the synthesized netlist of the design. The .spef file contains the RC

information of cell-to-cell routings which is extracted by Encounter. We perform Primetime statistical power analysis and timing analysis with the switching activity of both primary inputs and sequential outputs at 0.2. The area of the design is calculated by Encounter, and the die utilization ratio is set to be 0.6 which means 60% of the die area is used to place functional cells and the other 40% is used to place filler cells for providing extra routing space.

CHAPTER 4

ROUTABILITY IN S3DC vs. TR-L M3D

S3DC follows the mindset of the original Skybridge fabric that uses vertically-stacked gates interconnected in 3D on a template of vertical nanowires to yield orders of magnitude benefits over 2D CMOS. Core fabric aspects including device, circuit-style, connectivity, thermal management and pathway of manufacturing are co-architected for 3D compatibility. In this chapter, we will discuss the common routability issue in conventional 3D IC and how S3DC gets improved inter-cell routability.

4.1 Routability Issue in Conventional 3D IC

Conventional 3D ICs with die-die and layer-layer stacking using Through Silicon Vias (TSVs) [1] and Monolithic Inter-layer Vias (MIVs) [12] have been explored in recent years to generate circuits with considerable interconnect saving for continuing technology scaling. However, these 3D IC technologies still rely on conventional 2D CMOS's device, circuit and interconnect mindset showing only incremental benefits [12] while adding new challenges such as thermal management [13], manufacturing [14] and routability issues [4].

Among all conventional 3D IC approaches, transistor-level monolithic 3D IC (TR-L M3D) [16] represents the state-of-the-art that uses 3D standard cells for high-density design. But it still uses conventional via-to-metal routing structure as 2D CMOS where the standard cells are placed in a two-dimensional plane and routed by

the stacked metal layers above. This routing style provides limited routing capacity and routability to address high-density 3D routing which causes severe routing congestion [16] in large-scale circuits and diminishes its 3D benefits over 2D CMOS.

In the previous research [4] [16], the routing congestion rates of TR-L M3D in various benchmark circuits have been found. People tried to solve this issue by using standard cells with larger area and enhanced cell accessibility [4]. However, the overall 3D design density is thus reduced and degradation in design benefits against 2D CMOS are observed [4].

4.2 Routability in S3DC

S3DC is a fine-grained 3D IC fabric that uses vertically-stacked gates and 3D interconnections composed on vertical nanowires. This 3D fabric follows the mindset of our previous Skybridge fabric [5] that fully uses the vertical dimension instead of relying on a multi-layered 2D mindset. Its core fabric aspects including device, circuit style, interconnect are co-architected considering the common routability issue in 3D IC technology. In S3DC, the 3D interconnections provide greater routing capacity in both vertical and horizontal directions compared to conventional 3D ICs, which eliminates the routability issue while enabling ultra-high density design and significant benefits over conventional 3D ICs as well as 2D CMOS.

Compared with the conventional routing scheme, the S3DC's inter-cell and intra-cell routing has three main advantages: (i) input/output pins of each cell are place in multiple metal layers which realizes 3D routing access to the cell and thus

significantly improves the cell accessibility and inter-cell routability (See Fig. 3.3), (ii) each cell can have enough number of input/output ports to provide sufficient routing capacity for high-density 3D routing, (iii) the routing demand is evenly distributed in bottom-toup metal layers in contrast to TR-L M3D or 2D CMOS where the cell input/output pins are only placed in bottom metal layers resulting in busy routing and high congestion rate in these layers. These three factors contribute to significant reduction of routing congestion in high-density 3D design in S3DC. We carried out evaluation of routability in V3DC vs. the TR-L M3D by using both theoretical calculation and CAD simulation.

4.2.1 Theoretical Calculation using Rent's rule

The routability of 2D CMOS, TR-L M3D and S3DC are evaluated through analysis of routing congestion in benchmark circuits. Generally, the routing congestion in IC design is caused by the high-demand or over-demand of routing resource [33]. Thus, routing demand is a key metric used to reflect routing congestion and evaluate the routing complexity for a design before detailed routing [34]. We have done quantified evaluation for routing demand using the relationship between the routing demand *l* and the cell density *G* per unit area [34]:

$$l \sim G^{r-0.5} (r > 0.5) \qquad (1)$$

Where G represents the effective number of cells that need to be routed in a unit square and r is a constant known as the Rent's exponent [35]. The value of G can be calculated using Rent's rule [35] as shown in equation (2). Rent's rule is an empirical

observation about the relationship between the number of terminals (input/output pins) required by a design block to interface with its environment and the number of circuit components within the block [34]. It can be represented by the following equation:

$$E = A \cdot G^r => G = \left(\frac{E}{A}\right)^{\frac{1}{r}} \qquad (2)$$

where E is the number of terminals (input/output pins) in a unit square, A is the average number of terminals per cell. We assume all gates are distributed uniformly in the post-routed benchmark circuit. The parameter A is set to be 3 for each technology. The Rent's exponent r is set to 0.75 which is a typical value for large-scale designs [34]. Further, we use the pin number per micrometer square (pin density) as the parameter E. For 2D CMOS and TR-L M3D, the pin density E is reported by Encounter after 2D placement for a certain design. For S3DC, the accessible pins of each cell are distributed in multiple metal layers. Therefore, we calculated the pin density of S3DC's design by the expression:

$$E_{S3DC} = \frac{\#\ of\ Pins}{N \cdot S} = \frac{\#\ of\ Pins}{S} \cdot \frac{1}{N} = E_{\mathrm{ENC}} \cdot \frac{1}{N} \qquad (3)$$

N is the number of layers that are used to put pin accesses in our S3DC standard cell design. Its value is 5. S is the footprint of the die. N·S thus reflects the effective die area that is used to place cell pins. E_{S3DC} denotes the real pin density in S3DC's design. E_{ENC} is the pin density that is reported by Encounter that considers the cell pins distributed in a 2D plane and calculates the pin density using the die footprint S. It can be seen that the S3DC's effective die area for placing pins is multiple of the die footprint since the pin accesses distribute in multiple metal layers while in 2D CMOS

or TR-L M3D's effective die area used to place pins is just equal to 1x of die footprint. This contributes to significant pin density reduction in S3DC's designs in comparison

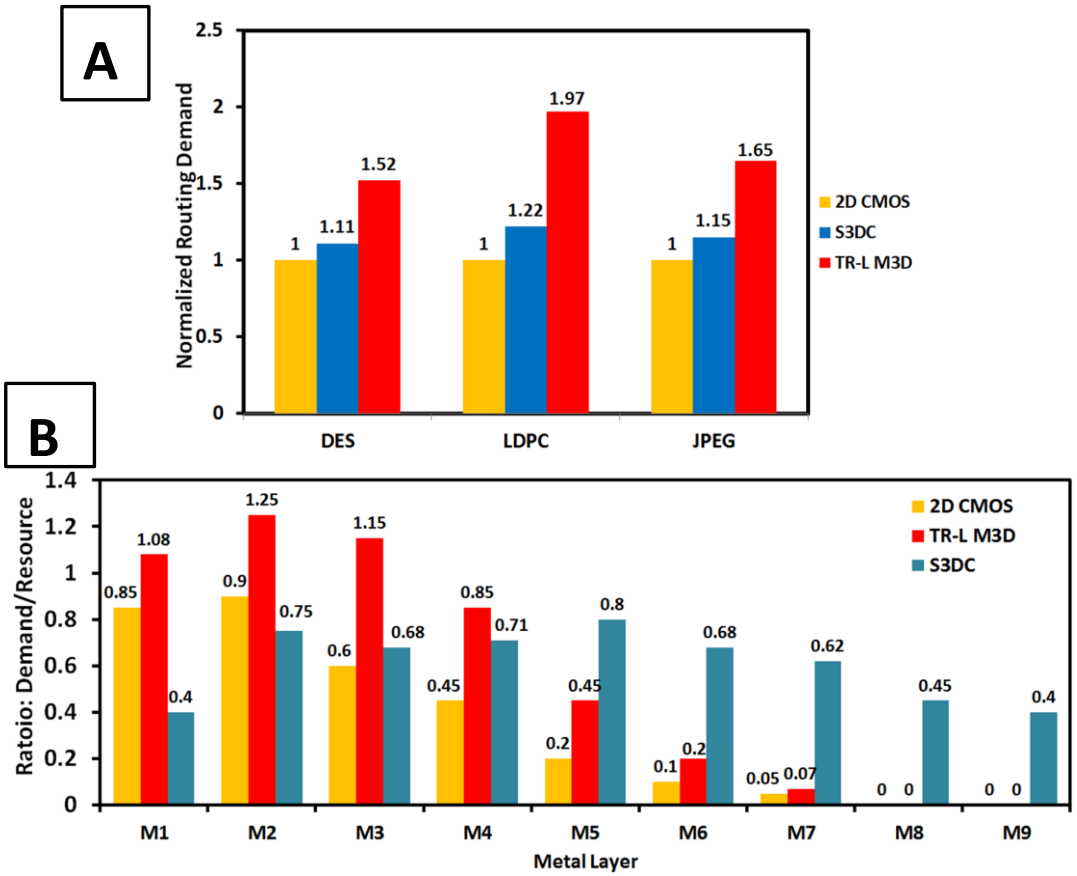

Figure 4.1 **A) Normalized routing demand in 2D CMOS, TR-L M3D, S3DC; B) Routing demand/resource ratio in all technologies' LDPCs**

to TR-L M3D which in turn significantly reduces the routing demand. Fig. 4.1A shows the normalized data of unit square's routing demand in each benchmark circuit for all technologies. The TR-L M3D'S DES and JPEG designs have around 1.6x routing demand over 2D CMOS while S3DC's designs have up to 15% increased routing demand compared to 2D CMOS. For the interconnect dominated core, LDPC, the TR-L M3D even shows 2x routing demand over 2D CMOS while S3DC has around 20% higher routing demand than 2D CMOS. It is also observed that S3DC has slightly higher routing demand over 2D CMOS while it has up to 1.6x lower routing

demand per unit square compared with TR-L M3D.

4.2.2 CAD-based Simulation

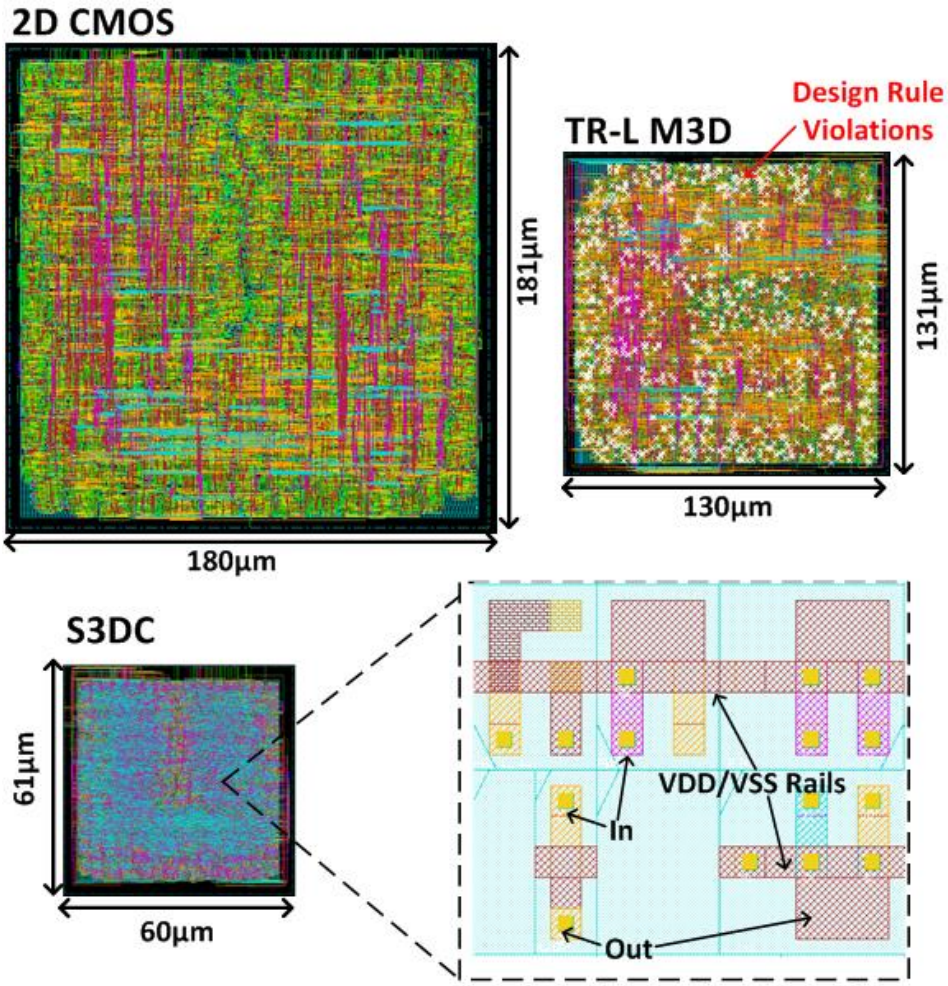

Figure 4.2 **Layouts of LDPC in 2D CMOS, TR-LM3D and S3DC**

By using the CAD flow shown in Section 2.1, we have done large-scale circuits benchmarking for S3DC, TR-L M3D and 2D CMOS. The Data Encryption Standard (DES), low-density parity-check (LDPC) and Joint Photographic Experts Group (JPEG) were chosen as benchmark circuits [36]. The design methodology in [16] is used for TR-L M3D's benchmark circuit design. Both 2D CMOS and TR-L M3D use the Nangate 15nm Library [25] as design PDK. Fig. 4.2 shows the layouts of LDPC

core design in 2D CMOS, TR-L M3D and S3DC, with clock tree, power delivery network, combination logic and sequential logic parts routed by Encounter. Due to high routing congestion rate, the TR-L M3D's design is routed with thousands design rule violations. By contrast, the S3DC's design is routed without any design rule violation while achieving 3x density benefit over 2D CMOS.

Fig. 4.1B shows the ratios of routing demand over routing resource in LDPC core design of each technology. These ratios that represent different metal layers are all reported by Encounter after layer-by-layer detailed routing. It can be observed that the TR-L M3D's LDPC design has over-demand routing in M1, M2 and M3 metal layers where the high-density routing for input/output pins are required. By contrast, for S3DC's LDPC design in Encounter, the routing demand distributes evenly in multiple metal layers with a maximum demand/resource ratio of 0.8. This even distribution helps in reducing opportunity of over-demand routing in the CAD design for S3DC.

4.2.3 Full-chip benchmarking (Logic+ Memory)

In addition to the logic parts deign, the memory parts in digital design needs to be considered and evaluated. The OpenSPARC T2 core [52] was used for evaluation. The OpenSPARC T2 core consists of 13 Function Unit Blocks (FUBs) including two integer execution units (EXU), a floating point and graphics unit (FGU), five instruction fetch units (IFU), and a load/store unit (LSU). Each FUB is synthesized with a 28nm cell library. In our implementation, top-level logic cells, i.e., cells outside FUBs, are grouped during synthesis to form an additional block. Thus, a total of 14

FUBs are floorplanned, and special cares are taken to use both connectivity and data flow between FUBs to minimize inter-block wirelength. Fig. 4.3 shows the memory placement and completed routing in Cadence Encounter [32]. The S3DC based SPARC T2 Core design was completed without any design rule violations. The library files of S3DC memory and M3D are scaled from the original 2D version provided by PDK. These libraries files with the standard cells' library files (See Chapter 3.2) were imported into the ASIC CAD flow to produce the design.

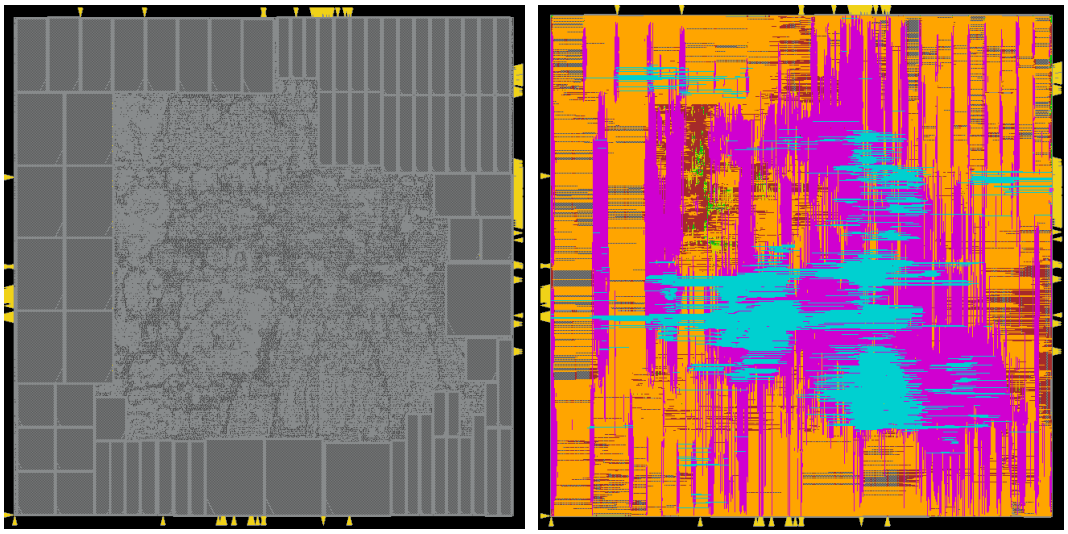

Figure 4.3 **Layouts of LDPC in 2D CMOS, TR-LM3D and S3DC**

4.3 Evaluation of Key Metrics

The key metrics of the benchmark circuits are evaluated to reflect the design benefits contributed by routability. The active power of each design is measured with uniform 1GHz clock frequency. And the area is reported by Encounter after placement. Table I shows evaluation results. The normalized footprint data shows that S3DC has up to 9x density against 2D CMOS, and the TR-L M3D has around 2x density. The reduction of routing demand in conjunction with compact vertical 3D gate design

contribute to about 3.3x shorter cell-to-cell wirelength which achieves up to 2.5x lower power against 2D while the TR-L M3D only has up to 1.4x shorter wirelength and 1.25x wire power efficiency. Since the VGAA transistor has much lower parasitic capacitance than conventional Finfet with junction [15], our S3DC's standard cells

Table 4.1: Results of Benchmarking

Benchmark Name	Design Type	# of Trans.	Best Frequency (GHz)	Wire Power (mW)	Cell Pin Power (mW)	Cell Internal Power (mW)	Total Power (mW)	Footprint (µm x µm)
AES	2D	648K	4.6	15.4	30.1	41.2	86.8	305x305
	TR-L M3D		5.3 (+15%)	11.8 (-22%)	25.6 (-15%)	34.6 (-17%)	72.0 (-17%)	220x220 (-51%)
	S3DC		4.1 (-12%)	7.4 (-52%)	5.1 (-83%)	14.8 (-64%)	27.1 (-69%)	110x110 (-87%)
LDPC	2D	300K	1.9	53.6	31.1	43.8	128.6	240x240
	TR-L M3D		2.2 (+17%)	40.7 (-24%)	26.1 (-16%)	35.5 (-19%)	102.3 (-22%)	170x170 (-50%)
	S3DC		1.7 (-10%)	20.4 (-62%)	5.9 (-81%)	15.3 (-65%)	41.6 (-63%)	81x81 (-89%)
SPARC T2 Core	2D	3267K	1.2	68.8	42.9	117.8	229.5	585x585
	TR-L M3D		1.37 (+14%)	55.0 (-20%)	36.0 (-16%)	98.9 (-16%)	189.9 (-19%)	421x421 (-52%)
	S3DC		1.1 (-8%)	22.9 (-66%)	7.3 (-83%)	43.6 (-63%)	73.8 (-67%)	200x200 (-88%)

have much lower driving capacitance, which achieves 6x lower cell pin power. The compact 3D standard cell design contributes up to 3x cell internal power efficiency. For interconnect dominated core, LDPC, the S3DC has 2.5x total power efficiency in comparison to 2D CMOS while the TR-L M3D around 1.25x power efficiency. For the cell-dominated core, AES, the S3DC achieves up to 3x total power efficiency over 2D CMOS while the TR-L M3D has 1.2x lower power compared to 2D. S3DC has around 10% performance degradation compared with 2D CMOS due to the usage of VGAA transistors, which have higher-resistivity channels [15]. This performance disadvantage however, can be overcome in multi-million transistor designs due to better routablity and shorter wire lengths [8].

CHAPTER 5

POWER DELIVERY NETWORK DESIGN

Design for power-delivery network (PDN) is one of the major challenges in 3D IC technology. In the typical layer-by-layer stacked monolithic 3D (M3D) approaches, PDN has limited accessibility to the device layer away from power/ground source due to limited routability and routing resources in the vertical direction. This results in an incomplete and low-density PDN design and also severe IR-drop issue. Some improved M3D approaches try to enlarge design area to create additional vertical routing resources for robust and high-density PDN design. However, this leads to degradation of design density and in turn diminishes 3D design benefits.

5.1 PDN Design Issue in Conventional 3D IC

The design for power-delivery network (PDN) is one of the major challenges in M3D which is caused by the routability issue. Due to limited routing capacity in vertical direction, PDN on top metal layers has poor accessibility to the device layer away from the power source. This leads to severe IR-drop in this device layer. In gate-level (G-L) M3D IC [12], large number of MIVs need to be used in cell-to-cell communication between top- and bot-tier while limited number of MIVs are used in the PDN's vertical routing to the bot-tier. Therefore, taking some cell-to-cell routing resources for PDN routing or enlarging design area to add routing resource for PDN, is the only way to achieve a robust and high-density PDN design in G-L M3D [12]. In the typical version of transistor-level (TR-L) M3D [16], top-tier's high-density

routing creates blockages, which limit PDN's vertical routing access to bot-tier and results in an incomplete and low-density PDN design. In the improved TR-L M3D version, larger cell footprint is used to add additional vertical routing resource for PDN's access to bot-tier. Overall, in both G-L and TR-L M3D approaches, the insertion of a robust PDN design would impact 3D cell-to-cell routing density which in turn diminishes the benefits over 2D design.

Skybridge 3D CMOS (S3DC) [37] is a fine-grained 3D IC fabric that uses vertically-stacked gates and 3D interconnections composed on vertical nanowires to yield orders of magnitude benefits over 2D ICs. This 3D fabric fully uses the vertical dimension instead of relying on a multi-layered 2D mindset. Its core fabric aspects including device, circuit-style, interconnect and heat-extraction components are co-architected considering the major challenges in 3D IC technology. In S3DC, the 3D interconnections provide greater routing capacity in both vertical and horizontal directions compared to conventional 2D and 3D ICs [38], which enables its ultra-high density design and significant benefits over 2D. Also, the improved routing capacity in S3DC is beneficial for a robust and high-density PDN design whose presence would not impact or create blockages on the 3D cell-to-cell routing.

5.2 Robust PDN Design in S3DC

5.2.1 PDN Design and Major Issue in TR-L M3D

The PDN design in TR-L M3D follows the standard PDN design techniques which use topmost metal layers for global wires, one intermediate metal layer and

VDD/VSS rails in M1 (See Fig. 5.1A). First, the power and ground signals are fed

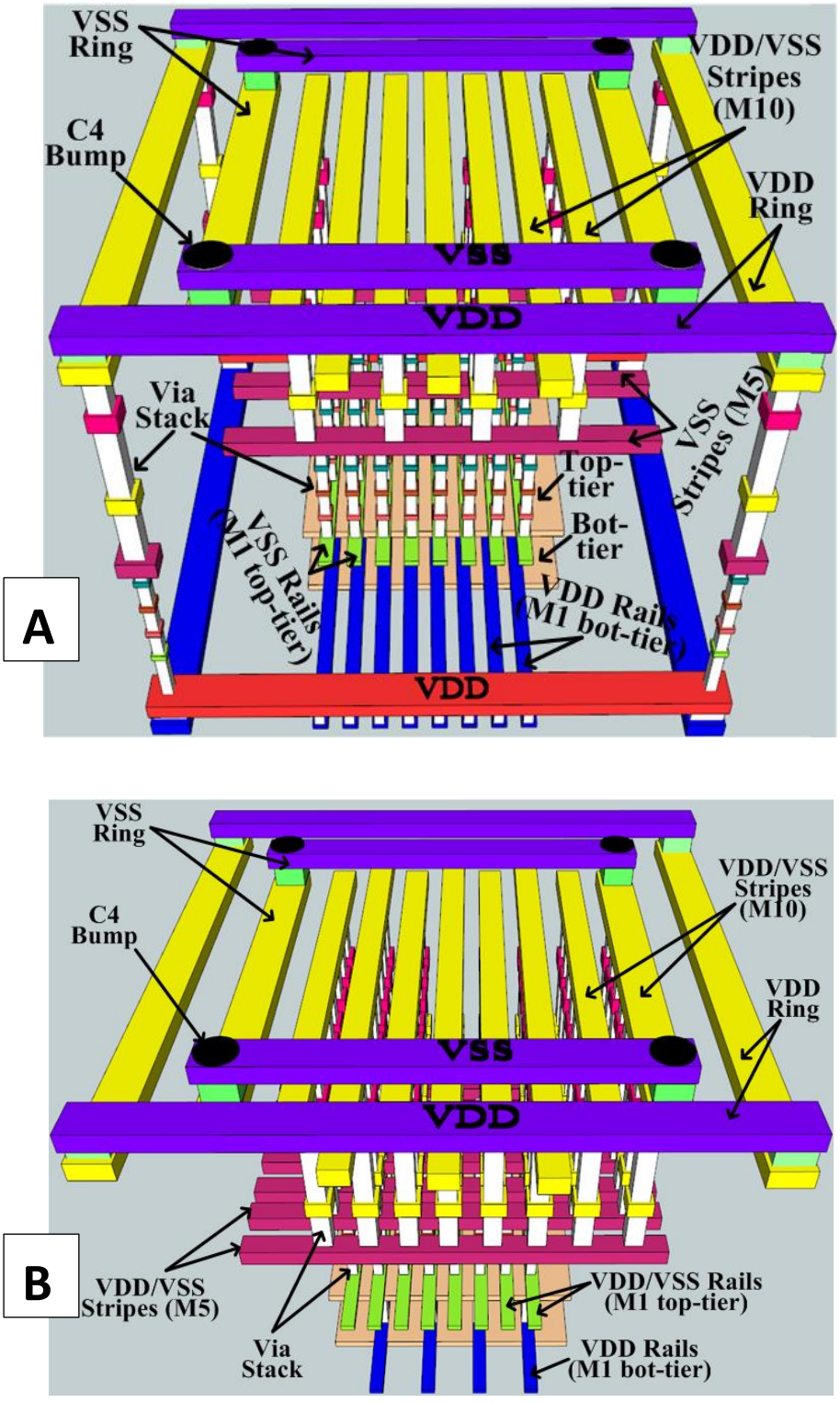

Figure 5.1 **A) Low-density PDN design in the typical TR-L M3D; B) High-density PDN design in the improved version of TR-L M3D**

from the C4 bumps to the VDD and VSS stripes in topmost metal layers (M10-11).

These power stripes also have ring connections at the periphery (See Fig. 5.1A) for lower resistance. Then, the VDD/VSS signals are delivered to the stripes in the intermediate metal layer (M5) by via stacks. These stripes have a finer pitch than the top metal layers (Fig. 5.1A). The stripes in the intermediate metal layer deliver VDD/VSS signals to local VDD/VSS rails that feed power to standard cells (Fig. 5.1A). In TR-L M3D, the local VSS and VDD rails are separated and placed into two tiers.

In the typical TR-L M3D approach [4][16], each standard cell is partitioned into two tiers; the pull-up network (PMOS) with its VDD rail is placed in bot-tier and the pull-down network (NMOS) with its VSS rail is placed in top-tier. The pull-up network exactly aligns with the pull-down network for optimal cell footprint shrinking. However, the VDD rails in bot-tier are thus blocked by the VSS rails in top-tier which leads to poor via accessibility to the VDD rails from intermediate metal layer in top-tier. Therefore, the typical TR-L M3D can only implement a low-density PDN design (See Fig. 5.1A) where VSS rails of cells are connected to its ground source by a network of high-density stripes and via stacks and VDD rails of cells are only connected to its power source by limited number of via stacks that directly connect the VDD rails to the rings at the periphery of the design block (See Fig. 5.1A). It is an intrinsic drawback in TR-L M3D that the top-tier's routing creates blockage on the vertical routing access to bot-tier, which in turn limits the communication between top- and bot-tier. In [4], the improved version of TR-L M3D uses larger cell footprint to provide additional vertical routing resource for access to the bot-tier. In this

approach, each 3D standard cell has both VSS and VDD rails in M1 of top-tier which can connect to VDD/VSS sources by standard PDN structure (See Fig. 5.1B). The VDD rails in bot-tier are aligned with the VDD rails in top-tier and connected by via stacks. It enables a high-density and robust PDN design where both VDD and VSS rails of cells are connected to their power/ground sources by a network of high-density stripes and via stacks. However, the major drawback is the footprint of 3D cell is increased due to the use of additional area for inserting VDD rails which impacts the design density and in turn diminishes the 3D benefits.

5.2.2 PDN Design in S3DC

S3DC fabric uses vertical nanowire based 3D gates for high-density 3D implementation instead of stacking multiple layers of 2D dies. As shown in Section II, stacking VGAA transistors and contacts on vertical nanowires enables a vertical cell design that has VDD rails on top metal layer M9 and VSS rail in M1. Therefore, the VDD rails in M9 can be easily connected to VDD stripes in top most metal layers (M10-M11) without using any intermediate metal layer. Also, the coaxial routing structure can provide significantly improved routability in vertical direction which enables high-density via connections between VSS rails in M1 and VSS stripes in the topmost metal layer. Fig. 5.2.A-B show the detailed PDN design in S3DC: the VDD/VSS stripes with rings are placed in M10-M11 which are added on top of the nanowire array and connected with C4 bumps; VDD rail (M9) of each standard cell is connected to VDD stripes (M10) using only one via layer; VSS signals are delivered

from VSS stripes in M10 to each VSS rail that are on the top (M9) of each routing

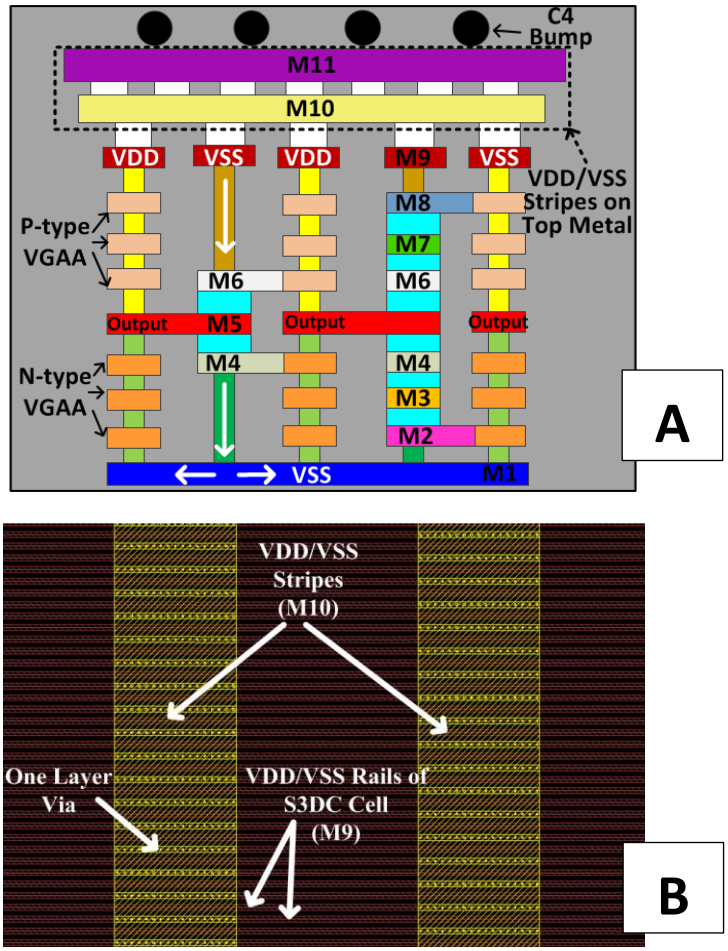

Figure 5.2 **A) PDN design in S3DC; B) S3DC's PDN routing implemented in** nanowire row; the routing nanowires deliver the VSS signals to the VSS rails of standard cells in M1. In this PDN design, the routing resources of M9/M1 and the vertical routing nanowires (inner routing layer of coaxial routing structure) are fully used for PND routing. The horizontal routing resources of M2-M8 and the vertical routing resources provided by the outer metal shell layer of coaxial routing structure are used in cell-to-cell 3D routing. This way, the cell-to-cell routing and PDN routing are completely separated and have no routing impact or blockage to each other. Considerable vertical routing resources can thus be used to design a robust and high-density PDN.

5.2.3 Methodology of PDN Extraction and IR-Drop Evaluation

Detailed IR-drop analysis was performed in large-scale benchmark circuits. The gate-dominated design AES and interconnect-dominated design LDPC were chosen for benchmarking. The benchmark circuits are implemented in both TR-L M3D and S3DC with uniform 16nm technology node. For both TR-L M3D and S3DC, the design and analysis use commercial CAD tools and encompass all steps from device characterization, RTL synthesis, PDN design, cell placement and routing, to system-level IR-drop evaluation.

The design of S3DC uses the device-to-circuit CAD flow published in [39]. First, we prepared basic design kit of S3DC that includes detailed effects of material choices, confined dimensions, nanoscale device physics, and associated 3D interconnect design rules and RC extraction table. Then the standard ASIC design flow was performed to generate the PDN designs for the benchmark circuits. In this step, the PDN design just includes the VDD/VSS paths from stripes in M10/M11 to the rails in M9. The VSS delivery paths (from M9 to M1) through silicided vertical nanowires were not implemented in this step since the design tool is not able to implement the coaxial routing structure that contains two layers of vertical routing. In the CAD design stage, only the outer metal shell layer of the coaxial structure was implemented by the vertical via stack between M1 and M9 and used in the cell-to-cell routing. The inner layer of coaxial routing structure (silicided vertical nanowire) which is used for the VSS delivery path from M9 to M1 is not included in the design stage but will be later added into the parasitic extraction results after the design stage

in order to capture the full design that contains both inner and outer routing layers. We then performed Sentaurus TCAD [20] to capture the series resistance of the silicided p-type nanowire, inter-layer contact structure and silicided n-type nanowire in a vertical routing nanowire (See Fig. 5.3). We directly added this resistance value into the extraction results of each VSS delivery paths after the parasitic extraction stage of the full design, since in S3DC adding the PDN routing would not change designed cell-to-cell routings. This way, the updated extraction results can fully capture the parasitics of the S3DC design that has cell-to-cell routing and PDN routing in parallel in the coaxial routing structure. At last, we performed static IR-drop analysis based on the extracted results using Cadence Voltus [39].

The methodology in [16] was used in the design of TR-L M3D. First, design kit was prepared based on a modified Nangate15nm PDK [25]. As discussed in Chapter 5.2.4, the TR-L M3D with low-density PDN uses different 3D cell structure compared to TR-L M3D with high-density M3D. We created 3D cell library versions for both TR-L M3D approaches. Next, the ASIC flow shown in [16] was used to encompass all steps of benchmarking from RTL to GDS layout. The design was then extracted for IR-drop analysis in Cadence Voltus [39]. Also, we performed IR-drop analysis for PDN design in 2D CMOS using Nangate 15nm PDK [25]. The PDN designs in TR-L M3D and 2D CMOS use the same density of VSS/VDD power stripes in intermediate layer (M5) and topmost metal layers (M10-M11) for fair comparison. The pitch and placement of C4 bumps follow the standard design rules shown in [40].

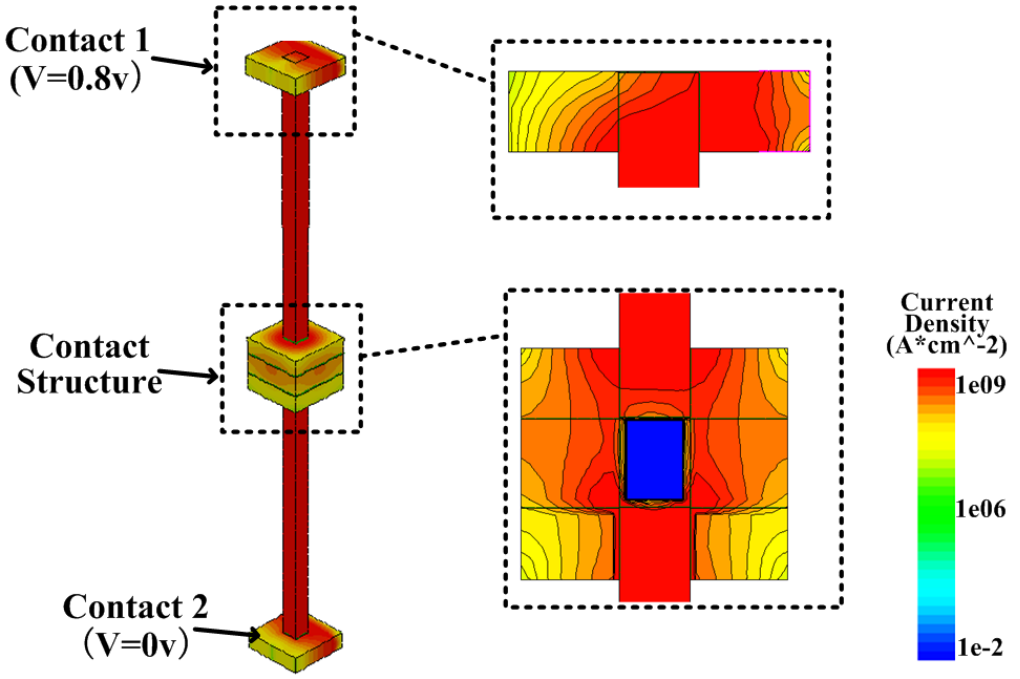

Figure 5.3 **Current density distribution in Sentaurus TCAD simulation of silicided vertical routing nanowire**

5.2.4 IR-drop Distribution in S3DC vs. TR-L M3D

5.4A-C shows the VDD IR-drop distribution of AES benchmark in TR-L M3D and S3DC. S3DC even has better IR-drop compared to the TR-L M3D with high-density PDN which is attributed to S3DC's significant routing resource that used in the PDN design.

Table II shows the average IR-drop in both LDPC and AES benchmarks. For VSS signal, both TR-L M3D and S3DC are within standard IR-drop budget (<5%*VDD). For VDD signal, the TR-L M3D with low-density PDN is out of standard IR-drop budget. TR-L M3D with high-density PDN has no IR-drop issue in VDD signal; it shows a 3x lower VDD IR-drop in LDPC and a 2.5x lower VDD IR-drop in AES compared to the TR-L M3D with low-density PDN. S3DC even shows 3x lower VDD drop in LDPC and 2.6x lower VDD drop in AES compared to TR-L M3D with

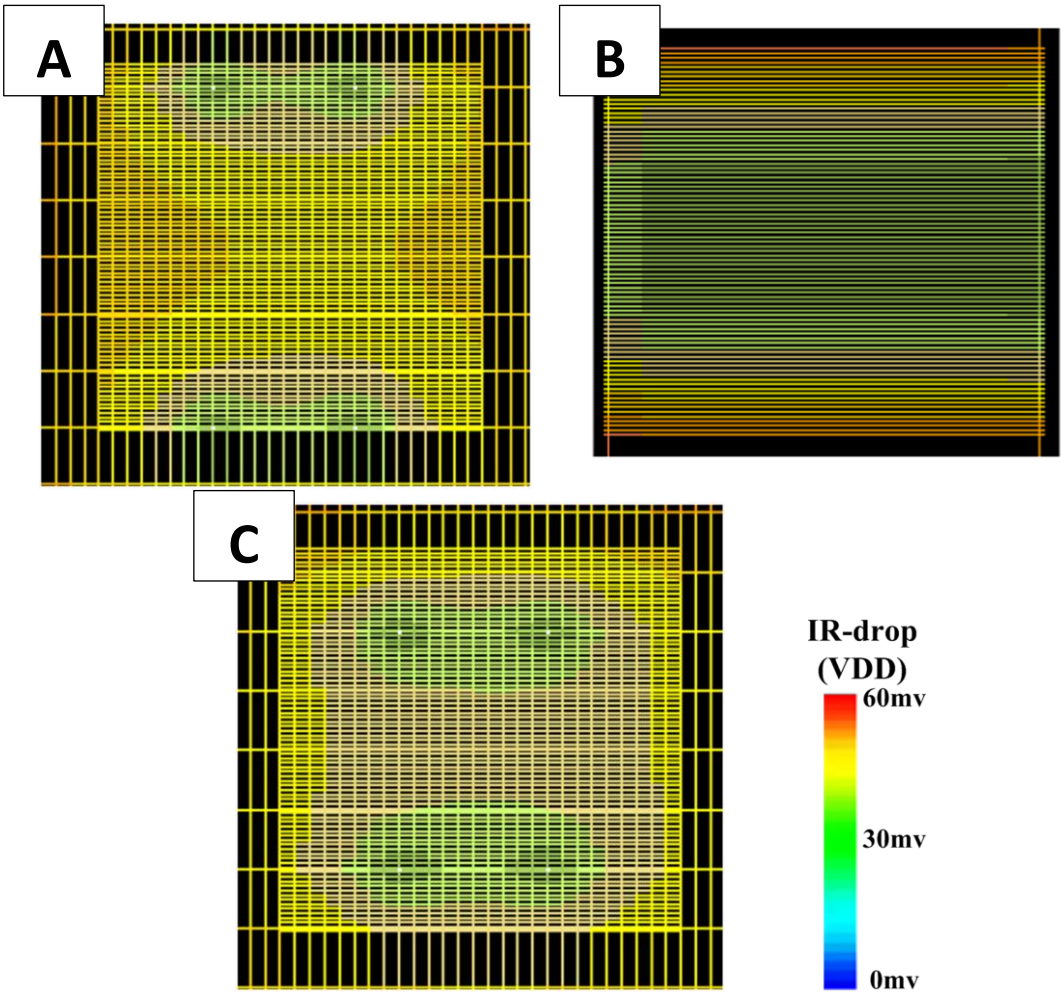

Figure 5.4 **IR-drop distribution in AES benchmark simulated in Cadence Voltus: A) Top-tier of TR-L M3D with high density PDN; B) Bot-tier of TR-L M3D with high density PDN; C) S3DC**

high-density PDN. Overall, both TR-L M3D with high-density PDN and S3DC can meet the requirement of standard IR-drop budget. Also, it can be observed that AES benchmark always has larger IR-drop compared to the LDPC benchmark. This is caused by the huge number of cells in AES core which leads to large total current flowing through PDN. However, an S3DC cell has significantly reduced cell parasitics [38], which results in cell power efficiency followed by total current reduction. This is a secondary factor that contributes to S3DC's lower IR-drop in comparison to TR-L M3D as well as 2D.

Table 5.1: Average IR-drop (Unit: mv)

Technology	LDPC (VDD=0.8v)		AES (VDD=0.8v)	
	VDD	VSS	VDD	VSS
2D CMOS	22	27	32	38
TR-L M3D (low-density PDN)	62	21	78	32
TR-L M3D (high-density PDN)	21	23	31	34
S3DC	7	14	12	18

5.3 PDN's Impact on Routing Congestion

5.3.1 PDN's Impact on Routing Congestion

In conventional 2D CMOS technology, the presence of PDN creates certain routing blockages on cell-to-cell routing (cell-to-cell routing is designed after PDN design). Therefore, in conventional 2D design, the trade-off between PDN robustness and cell-to-cell routing efficiency needs to be carefully addressed. In M3D ICs, the cell-to-cell routing has higher (2x) routing density than 2D CMOS, which means the insertion of PDN results in more blockages and heavier congestion on cell-to-cell routing. This would easily lead to a non-optimal design which has severely increased total wire length and caused degradation of 3D design benefits. Fig. 5.5 shows the routing of M2, M4, M5 and M6 in the AES benchmark of TR-L M3D with and without PDN (low density PDN). It can be observed that the presence of VDD/VSS stripes in M5 leads to extreme busy routing in M5. The cell-to-cell routing in M6 also becomes much denser due to the heavy routing congestion in M5. Additionally, the presence of via stacks (V1-V5) of PDN creates severe blockage and results in denser routing in M2 and M4 compared to the design without PDN. In the TR-L M3D with

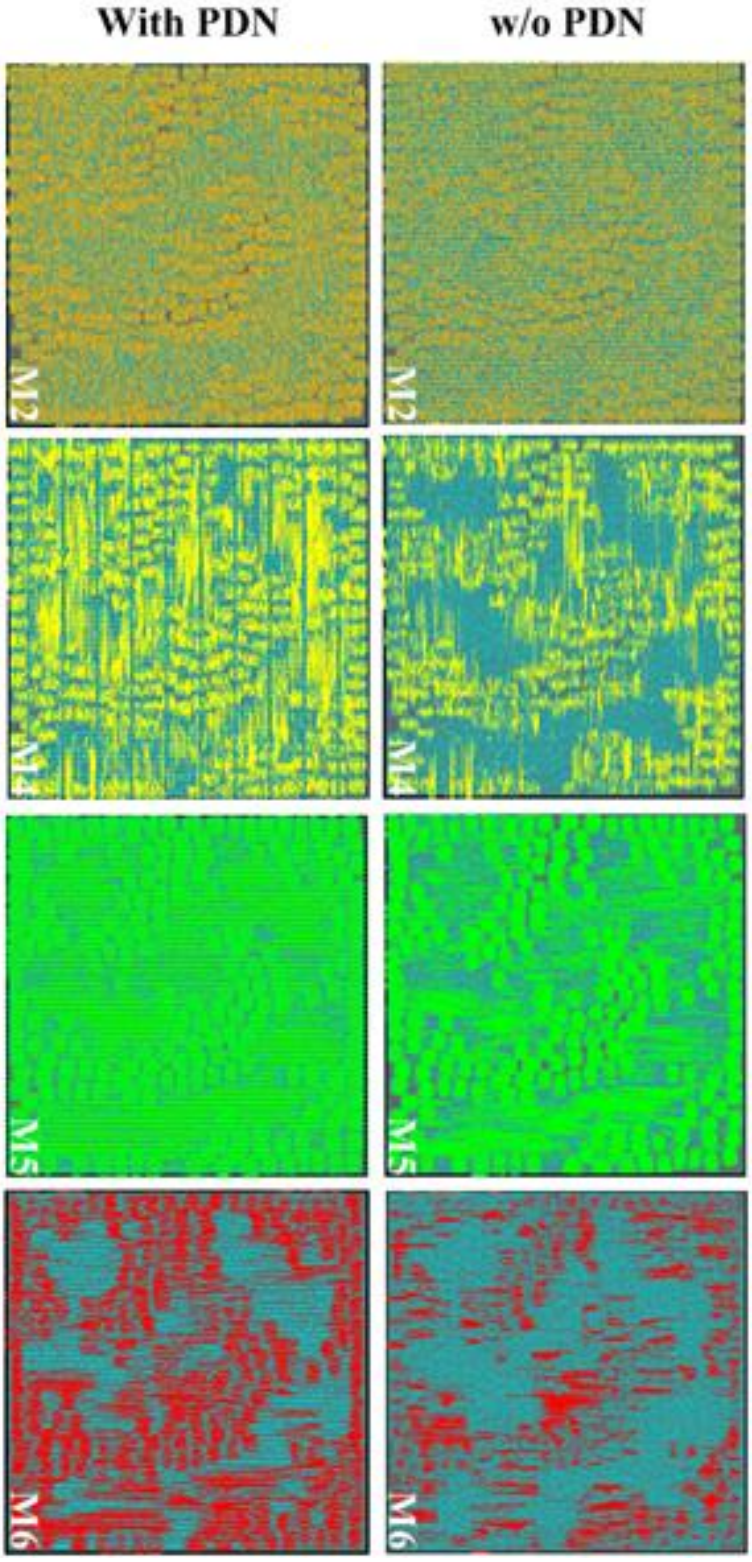

Figure 5.5 **Routing congestion comparison of AES benchmark of TR-L M3D with and without PDN (low-density PND version)**

high-density PDN, the PDN routing would have more impact on cell-to-cell routing.

In S3DC, the coaxial routing structure can provide 2 layers of vertical routings (See

Fig. 2.2C); the PDN uses the inner layer (silicided nanowire) and the cell-to-cell routing uses the outer layer (the metal shell around a nanowire). This way, the PDN routings are completely separated from cell-to-cell routing and have no routing blockage on cell-to-cell routing. Thus, in S3DC the PDN insertion has on impact on 3D cell-to-cell routing. Also, sufficient routing resource can thus be provided for a robust and high-density PDN design that meets the requirement of the standard IR-drop budget.

5.3.2 PDN's impact on Signal Integrity

As shown in last section, the insertion of PDN has severe impact on routing density especially in the designs of M3D. Even without the PND insertion, the routing density in M3D is much larger than 2D CMOS due to significant improvement in design density. Fig. 5.6 shows the M2 routing density in the LDPC benchmark. It can be easily observed that the M2 layer's routing in M3D is much busier than the M2 layer routing in 2D CMOS. Similarly, it would happen in S3DC because S3DC enables around 9x design density benefits compared to 2D CMOS (See Chapter 4.3). Also, as discussed in Chapter 4, the average wirelenght is reduced in M3D and S3DC compared to 2D CMOS. Then the coming question is how's the signal integrity in S3DC and M3D compared to 2D CMOS. This question is based on the consideration that the increased routing density will absolutely increase the cross-talk between the signal nets while on the other hand the reduction in average wirelenght will improve the signal integrity.

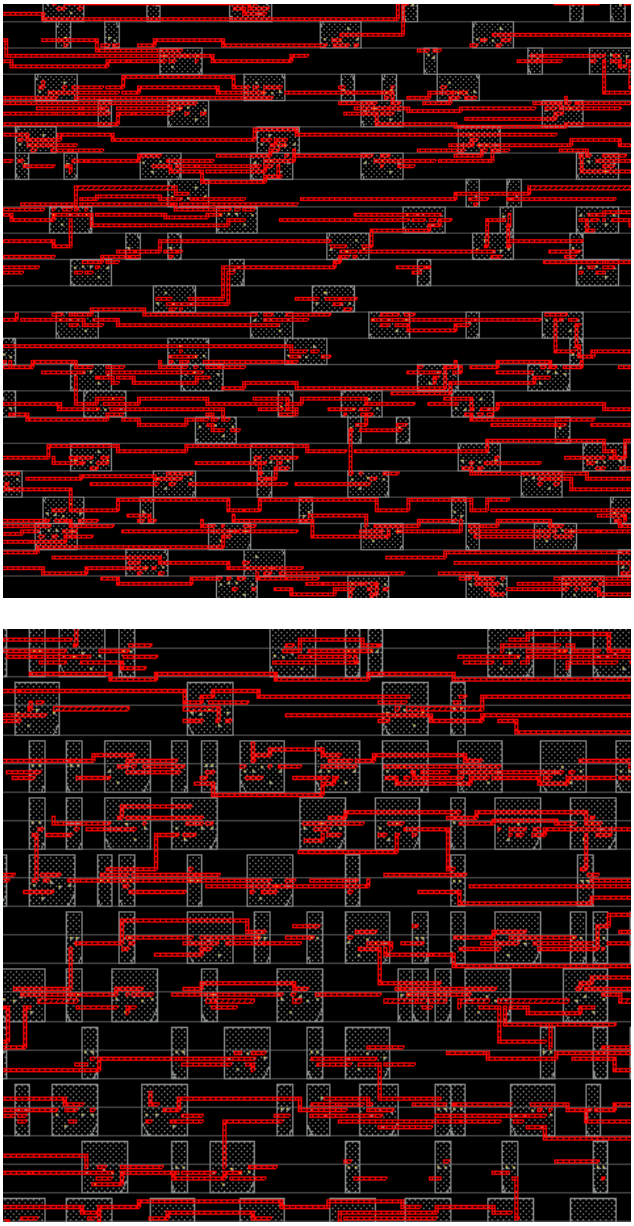

Figure 5.6 **A) M2 Density in 5µm* 5µm Square (TR-L M3D); B) M2 Density in 5µm* 5µm Square (2D CMOS)**

A SI evaluation methodology was developed to full evaluate the SI degree in each technology based designs. Fig. 5.7 shows the design and evaluation flow. Firstly, the ASIC design flow and PDK were used to generate finish the LDPC benchmark design with imported Verilog codes. Then, based on the generated SEPF file [32] which contains the parasitic information of each net and was dumped from Cadence Encounter [32], a Perl scripted was developed to extract the RCs of each net

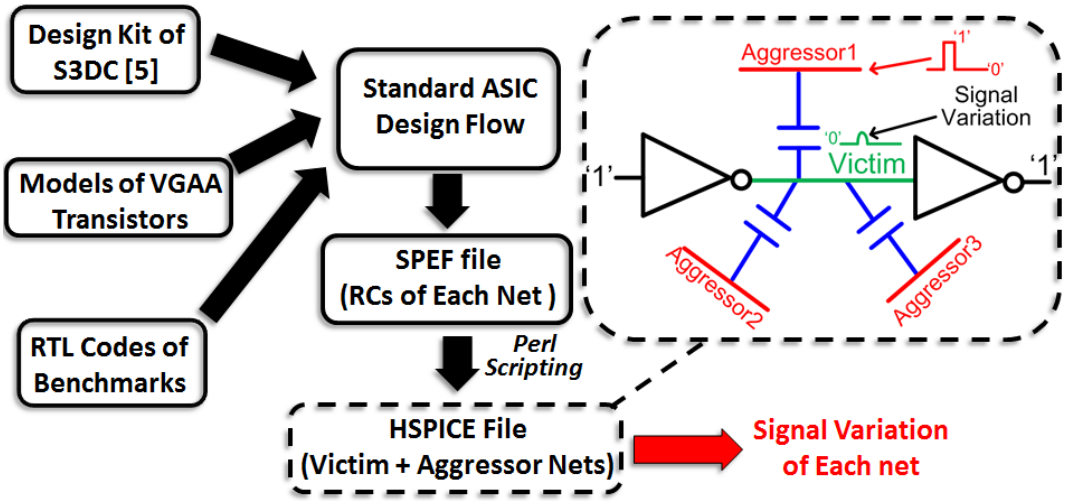

Figure 5.7 **Methodology of SI analysis**

and write it into a HSPICE-compatible format. In the generated HSPICE simulation file, each net contains the full RC information from the SPEF file and was driven by a inverter with standard sizing. Also, each net drives another inverter which forms as a load to the net. Each net is connected to aggressor signals (an ideal pulse) through the coupling capacitances which are extracted based on the SPEF file. After the simulation in HSPICE, the peak voltage of each net's signal variation was fully evaluated by HSPICE and reported.

Fig. 5.8 shows the SI evaluation results. Compared to 2D CMOS the SI in TR-L M3D based LDCP design has better SI degree and impact. In the data, we only count the nets that have SI variation over 10μV. The LDPC design has totally 63K of nets. Through the comparison, it can be also observed that in both 2D CMOS and TR-L M3D, the insertion of PDN leads to degradation of SI in their LDPC designs. However, the SI degree in S3DC based LDPC design shows significantly improved SI degree compared TR-L M3D as well as 2D CMOS. Also, the insertion of PDN in S3DC helps in the improvement of SI degree.

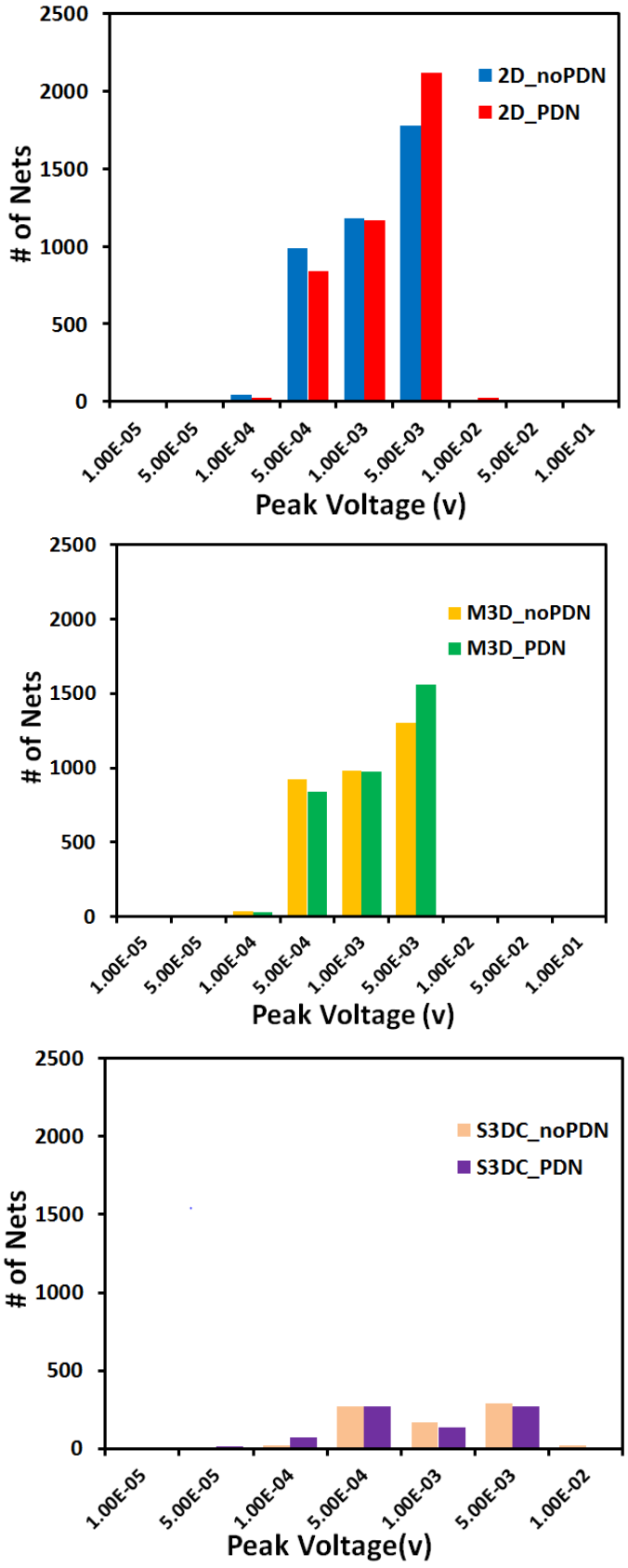

Figure 5.8 **A) SI Results in 2D CMOS; B) SI Results in TR-L M3D; C) SI Results in S3DC**

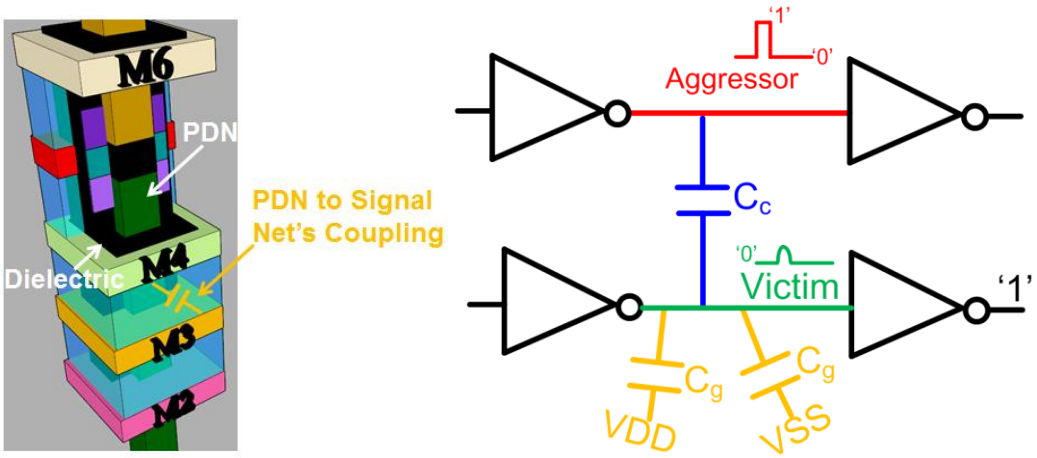

Figure 5.9 **PDN's help in SI improvement in S3DC**

The followed question is why the insertion of PDN in S3DC improves SI degree while the PDN in M3D and 2D CMOS worsen the SI degree. This is because that in S3DC the inserted PDN has no impact on signal routing and also adds gourd capacitance to each net to help in the improvement of SI. The PDN insertions in M3D and 2D CMOS also add ground capacitance to the nets that can improve SI degree. However, the insertion of PDN also leads to larger routing density (See Fig. 5.9) which increases the coupling capacitance between the nets and cancels the SI improvement from the added ground capacitance.

CHAPTER 6

SRAM DESIGN AND VARIATION TOLERANCE

The continuous push for denser, faster and more power efficient computing is driving CMOS scaling to its limit. Numerous new challenges are emerging related to power consumption, circuit noise, manufacturability and cost. These challenges are especially critical for CMOS SRAM circuits, where both PMOS and NMOS transistors need to be precisely sized and doped for memory operation and for sufficient noise margin. Due to the complex and compact layout of SRAM circuits, it is becoming difficult to maintain such precision at nanoscale. Moreover, controlling passive power in SRAM circuits is becoming a big concern; this is mainly because of the static SRAM circuit style and leakage current increase in nanoscale transistors.

6.1 Design and Scaling Issues in Conventional 3D SRAM

3D integration is an effective approach to reduce the chip footprint and increase the density. However, conventional TSV-based 3D technology [1] is proved not suitable for 3D SRAM cell because of the prohibitively large TSV size. On the other hand, M3D approach is used to enable such tighter alignment precision of the strata and the nano-scale inter-tier vias offering unparalleled opportunities for ultra-high density 3D SRAM compared with TSV-based approach. Currently, the M3D-baed SRAMS has been designed and extensively evaluated [43]. However, as mentioned, M3D itself still relies on via-to-metal interconnect mindset showing only incremental benefits in SRAM design. Additionally, the M3D SRAM inherits the short-channel effects and

sub-threshold slope degradation (SS) in technology scaling. Therefore, M3D-based SRAM is vulnerable to some critical variation sources like random dopant fluctuation, line edge roughness, line width variation as technology node scaling down. Also, this is a common situation in both M3D and 2D CMOS due to the use of the same transistor. In [43] and [44], people proposed the vertical/lateral gate-all-around FET based SRAM design which can overcome negative effects in technology node scaling and also can enhance the immunity to process variation. However, this approach can

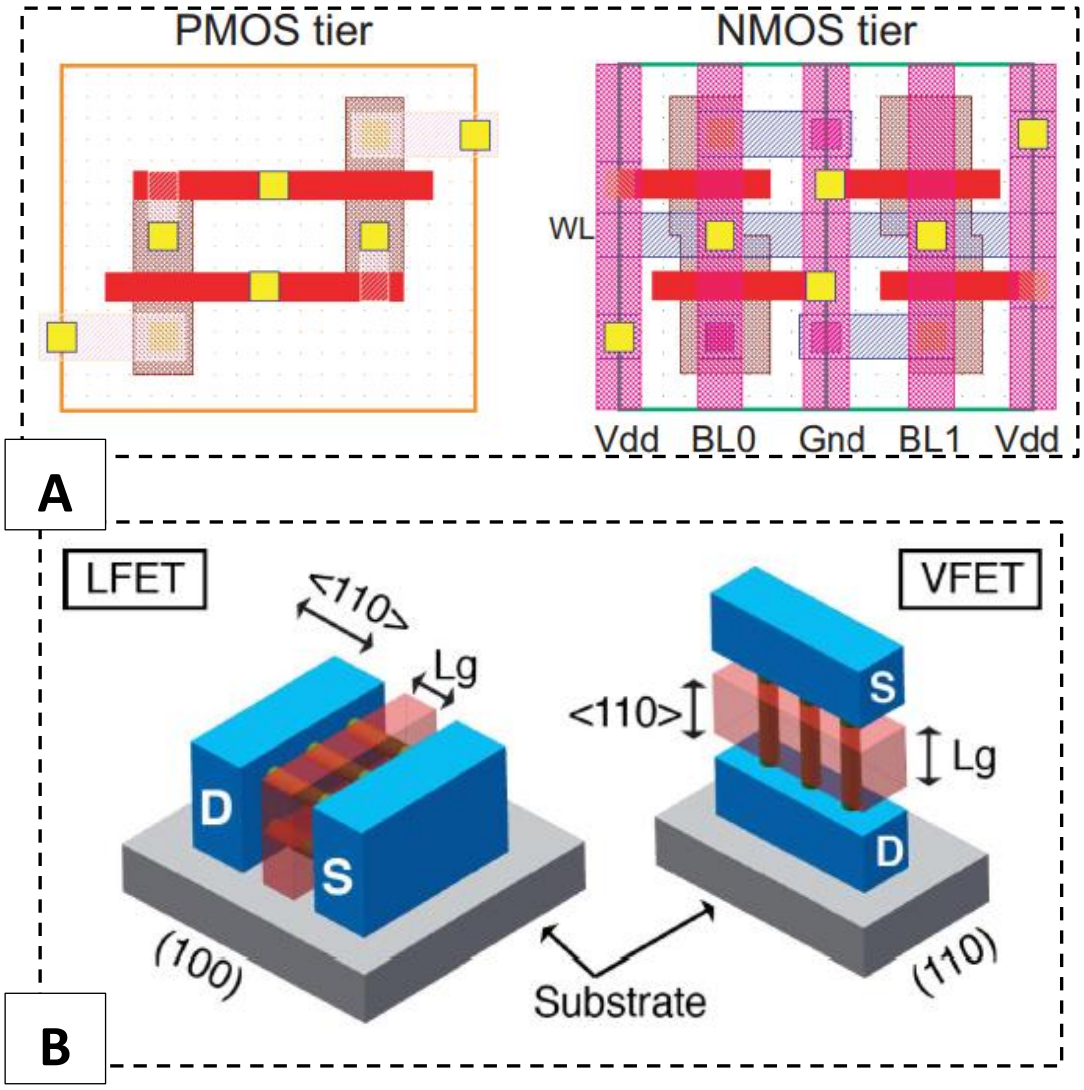

Figure 6.1 **A) Top-tier (PMOS) and bot-tier (NMOS) layout design in M3D-based SRAM [42]; B) Schematic views of a lateral stacked nanowire transistor (left) and a vertical nanowire transistor (right) [43]**

only improve the device but still follows the 2D SRAM design style. Therefore, in this technology the actual SRAM design benefits over 2D CMOS is incremental and the scaling factor is as small as 2D CMOS.

6.2 Design and Benefits in S3DC SRAM

6.2.1 SRAM Design in M3D

The M3D uses the same device and interconnect mindset as 2D CMOS and implements 3D design by stacking multiple 2D CMOS layers. In their typical approaches that stack two silicon tiers for 3D implementation [1][4] (our S3DC also uses two silicon layers), 2x design density [4] are achieved against 2D CMOS for each standard cell.

In [4][42], the M3D SRAM are designed by splitting PMOS and NMOS transistors into two tiers within a standard cell, and MIVs are used for cell internal vertical interconnection. Fig. 6.1A shows the layouts of a M3D SRAM cell. There are two metal layers (M1, M2), and one silicon layer (for PMOS) in the bottom tier (bot-tier) and one silicon layer in the top-tier (for NMOS), with an inter-layer-dielectric (ILD) for isolation. This way, the pull-up and pull-down networks of each inverter are splitted and that each silicon layer has only one type of transistor. The monolithic inter-layer via (MIV) penetrates the ILD and connects with M2 in the bottom tier and connects the pull-up and pull-down network of each standard cell.

For each standard cell in M3D, the number of PMOS is equal to the number of NMOS. This way, 50% footprint reduction can be achieved after splitting PMOS and

NMOS into two tiers and stacking them. However, in conventional 6T-SRAM design, 4 NMOS and 2PMOS are used in each cell. This unbalance leads to a reduced design benefits compared to other standard cell designs in M3D. Additionally, M3D still partially relies on conventional routing as 2D CMOS which limits the use of vertical dimension for design optimization. Therefore, in M3D SRAM [42] the length of bitline maintains the same as the conventional design in 2D CMOS and only the length of wordline gets reduced.

6.2.2 SRAM Design in S3DC

The SRAM design in S3DC follows the conventional 6T-SRAM design. A full use of vertical dimension is achieved by vertically stacking transistors and interconnect alone nanowires which helps in achieving a compact SRAM design. Detailed design is shown in the Fig. 6.2. The S3DC SRAM uses 9 metal layers and 6 nanowires: VDD rail is placed in M9, VSS rail is placed in M1 and routed to M9 rail through the internal layer (silicided nanowire) of coaxial routing structure, the bit line is placed in

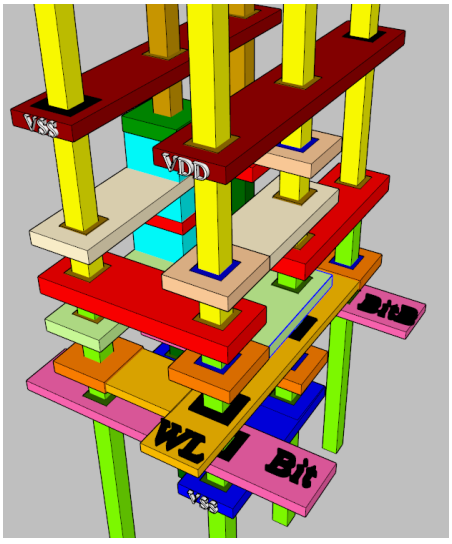

Figure 6.2 **Layout of SRAM design in S3DC**

M2, word line is placed in M3, other intra-cell interconnect are placed in M4 to M8. This design fully uses the vertical space and the significant routability in S3DC. The compact design results in significant bit line length reduction (around 60%) compared to 2D CMOS based SRAM design.

6.2.3 Evaluation of Key Metrics

In this work, the SRAM design in M3D, S3DC and 2D CMOS are all evaluated with a uniform 16nm technology node. The 2D CMOS and M3D SRAM design

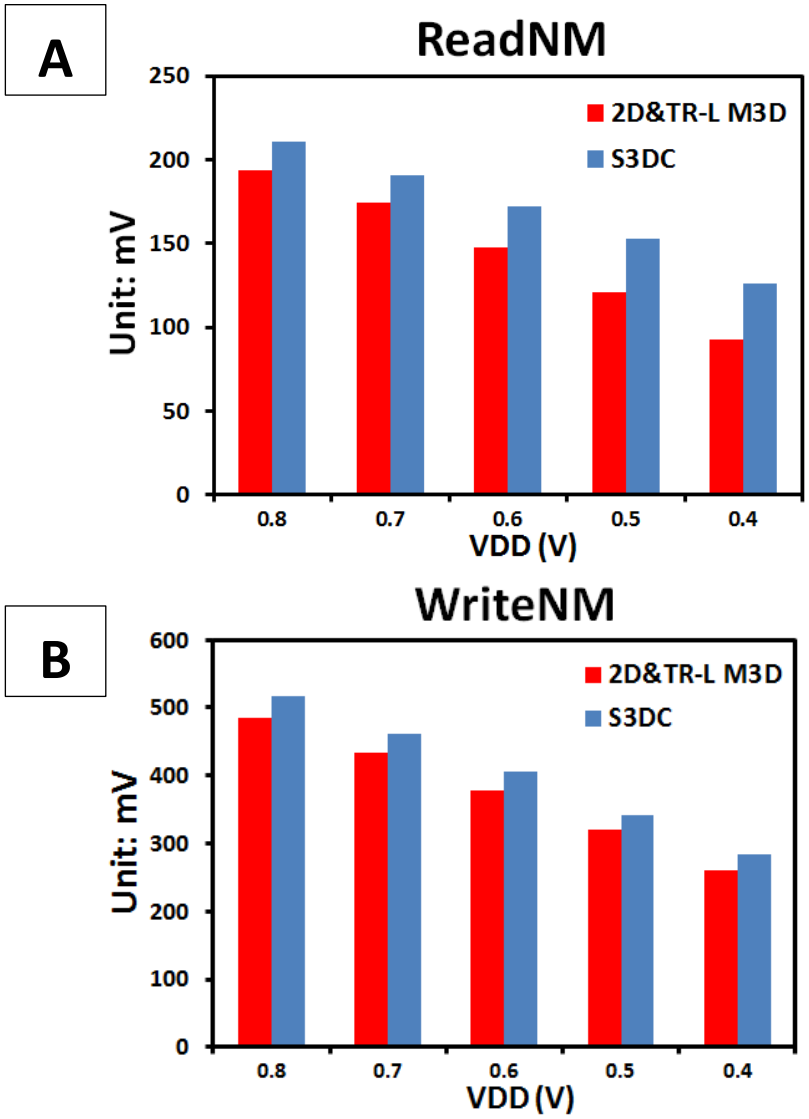

Figure 6.3 **Read and Write NM of each technology based 6T-SRAM**

follows the 1:1:1 cell design shown in [42][43]. The design uses the Nangate 15nm PDK to produce SRAM cell layout in Cadence Virtuoso [55] which was followed by Design Rule Check (DRC), Layout VS Schematic (LVS) for design validation. After layout design of SRAM, the RCs of SRAM are manually extracted using the Predictive Technology Interconnect Models [31], following the dimensions and material types of the structures in the layouts. Physical HSPICE netlists were then built following the circuit topology and the extracted RC. The M3D RCs were extracted by using the methodology in [4]. The 2D CMOS based SRAM was also benchmarked using conventional design and RC extraction tool. Both 2D CMOS and M3D uses the PTM 15nm PDK [25]. The simulation also assumes the SRAM cell is in a 32*32 array where the practical impact from the big capacitance of wordline and bitline can be included in the simulation.

Fig. 6.3 shows the comparison of read/write noise margin (NM) in S3DC, TR-L M3D and 2D CMOS based 6T-SRAM. The S3DC SRAM shows around 10%-20% better read/write NM with M3D and 2D CMOS. For the NM, as VDD scaling down, S3DC shows increased benefits due to lower Drain Induced Barrier Lowering (BIDL) in GAA transistor which results in significant V_{th} change as V_{ds} changes.

Fig. 6.4 shows the comparison of read/write time in each technology. The S3DC shows around 1.8x faster read over 2D CMOS and 1.3x faster read compared to TR-L M3D due to significant reduction of bit line. However, S3DC shows lower performance than TR-L M3D and 2D CMOS due to the higher resistivity channel in junctionless transistor compared to conventional junction transistor used in

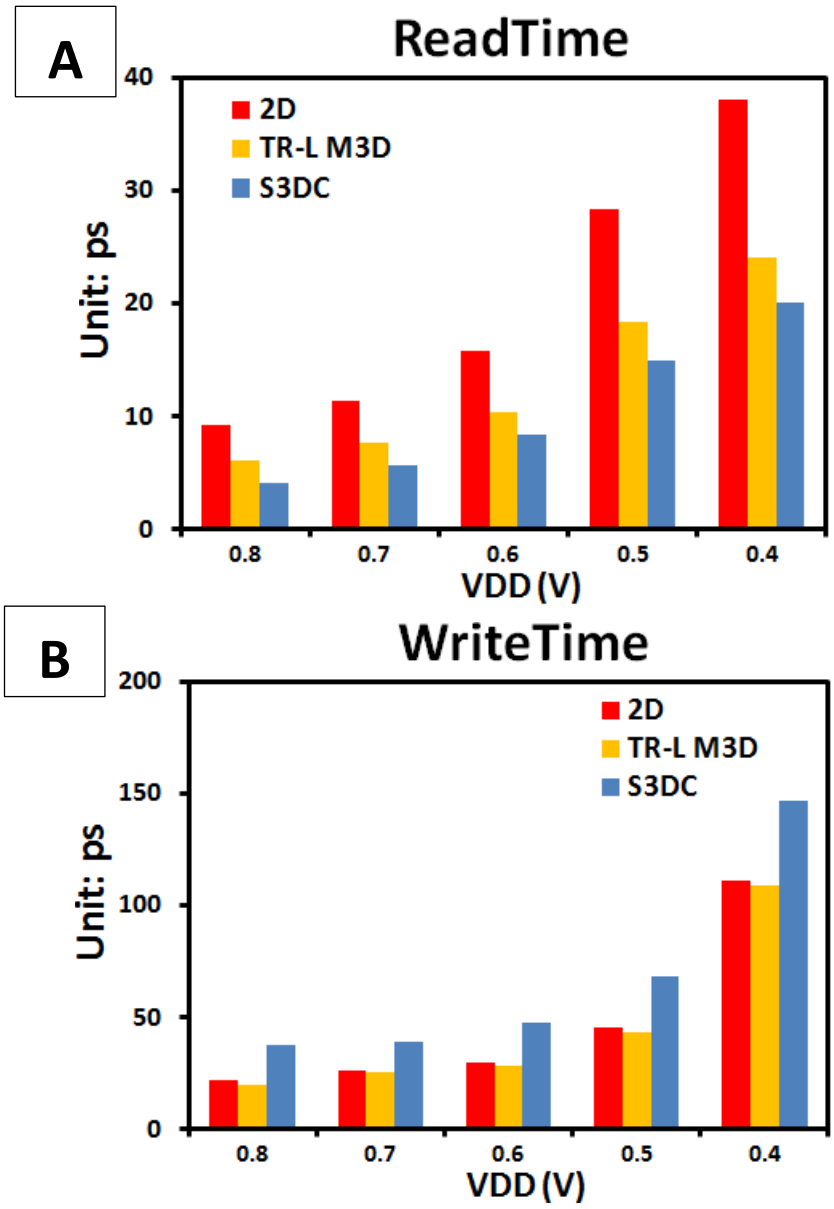

Figure 6.4 **Read and Write time of each technology based 6T-SRAM**

TR-L M3D and 2D CMOS. Fig. 6.5A shows the comparison of leakage. The S3DC SRAM has 8x lower leakage compared to TR-L M3D as well as 2D CMOS.

Additional step was done to validate the simulation results of S3DC SRAM. We used a TCAD-simulation based VGAA junctionless transistor model for S3DC SRAM evaluation with comparison to our analytical model shown in Chapter 6.3.1. Fig. 6.5B shows the results of read NM. Our analytical model based SRAM read NM can fully match with the read NM that simulated using TCAD-based model. This proves that

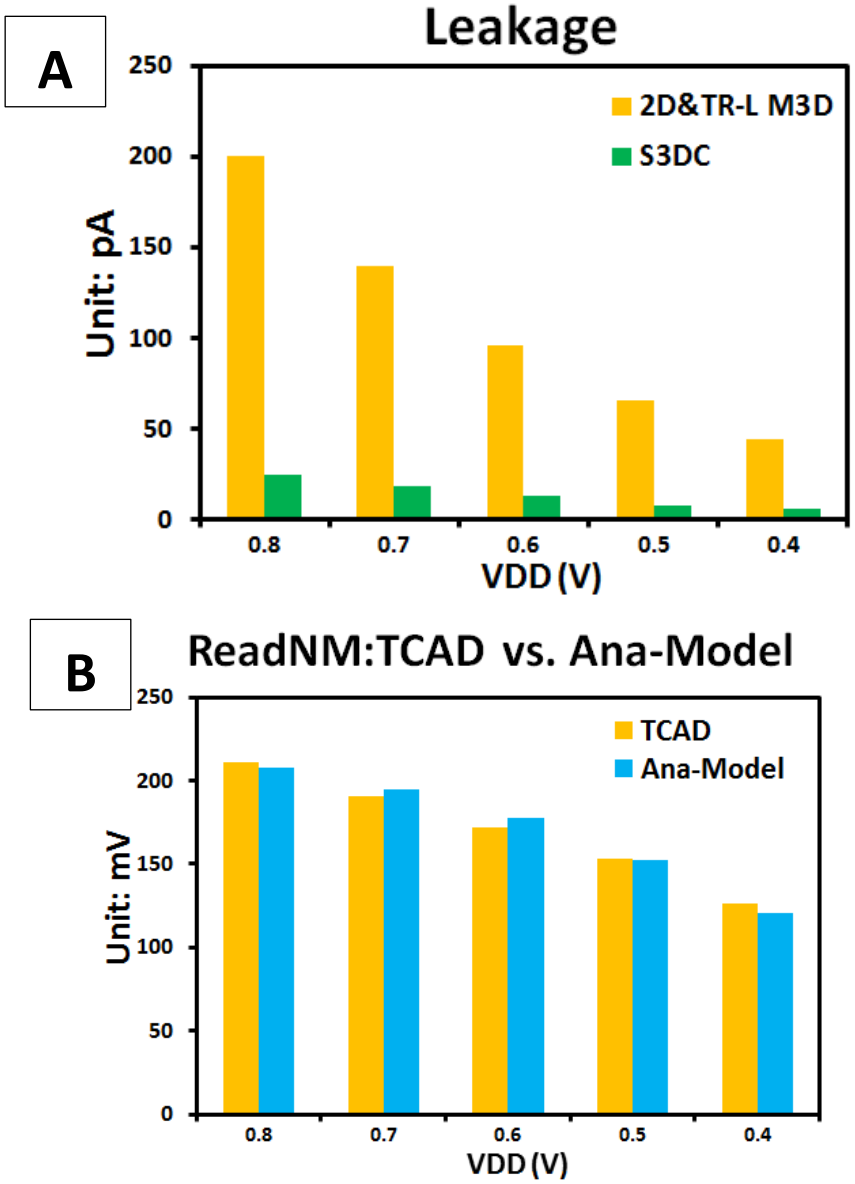

Figure 6.5 **A) Leakage of each technology based 6T-SRAM; B) Comparison for TCAD based model vs. our analytical model in read NM's evaluation**

our analytical device models can essentially capture the I-V and C-V characteristics of VGAA junctionless transistors and produce a precise evaluation of S3DC SRAM.

6.3 Variation Tolerance in S3DC SRAM

The ability of variation tolerance is a key metric in SRAM. It significantly impacts SRAM's failure rate. However, one of the major issues in technology scaling is the

increased short-channel effects which severely weaken device's immunity to process variation. In this section, we investigate the impact of variations in S3DC, M3D and 2D CMOS based SRAM designs. We consider the primary variation sources in the advanced technology node and evaluate the variations' impact on SRAM design's stability which is a key metric to reflect the failure rate of SRAM.

In the advanced 16nm technology node, the complicated FinFET device structure needs extreme precision in process and has strict requirement on variation control. This raises up various of variations sources that result in two kinds of impact on circuits (correlated and uncorrelated impacts). Variations that depend on particular process conditions, such as the uniformity of etching or annealing, or on particular aspects of the layout, such as the orientation or the proximity, will tend to systematically affect all devices or cells on a chip [49]. They can be modeled as random variables to account for process fluctuations; however, the high sensitivity to process or layout makes their distributions difficult to predict in a general analysis [49]. On the other hand, variations from uncorrelated random sources such as line edge roughness or line width variation are inherent to semiconductor processing and therefore more suitable for a general analysis.

It is well-known that SRAM is a symmetric cell which can be easily disturbed by mismatch of strengths between the two cross-coupled inverters. Therefore, the uncorrelated variations are the more significant cause of SRAM failure. The major impact from the uncorrelated variation is the change of Vth of each device which leads to significant change in voltage transfer characteristic (VTC) of each inverter.

This may severely squash the noise margin of SRAM and cause read/write failure. In the uncorrelated variations sources, the lithography line width variation is a more major variation source compared to ling-edge roughness due to its significant impact on channel length and width.

6.3.1 Analytical Model of VGAA Transistor

In this work, we focus on the lithography-caused line width variation that mainly results in channel length and width variation. In order to evaluate the impact of read NM from channel length and width, an analytical model of the VGAA transistor needs is carried out with accurate match with TCAD-based simulation results and compatibility with HSPICE simulation flow.

The device model is start from classic model and Boltzmann statistics. We can write the Poisson's equation in the silicon channel as:

$$\frac{d^2\varphi}{dx^2} = \frac{q}{\varepsilon_{Si}} N_D \left[\exp\left(\frac{\varphi - V}{v_T}\right) - 1\right] \qquad (1)$$

where q is the electronic charge, ε_{Si} is the permittivity of silicon, $v_T = kT/q$ is the thermal voltage, $\varphi(x)$ is the electrostatic potential, and V is the electron quasi-Fermi potential. Based on the boundary condition in junctionless transistor, the electrical field in the center of the channel should be 0. Equation (1) must satisfy the following boundary conditions:

$$\frac{d\,\varphi}{d\,x}\Big|_{x=0} = 0 \qquad \varphi(\pm\frac{t_{Si}}{2}) = \varphi_S \qquad (2)$$

Then we can generate the following equation that describes the relationship between

surface electrical field E_s and surface potential φ_s.

$$E_S = \frac{qN_D v_T}{\varepsilon_{Si}} [\exp\left(\frac{\varphi_s - V}{v_T}\right) - \exp\left(\frac{\varphi_0 - V}{v_T}\right) - \frac{\varphi_s - \varphi_0}{v_T}] \quad (3)$$

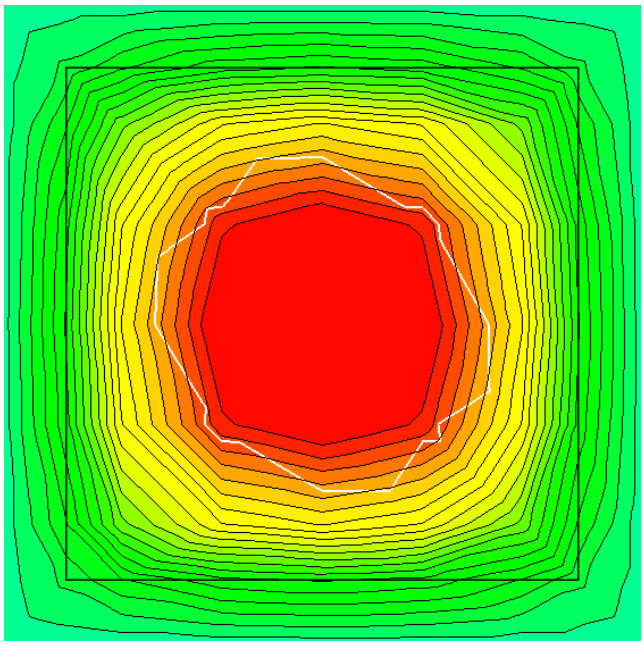

Figure 6.6 **Contour lines of electrostatic potentials inside channel**

Then, we assume the contour lines of electrostatic potentials inside channel follows a round profile. This assumption has also been used in the other models of junctionless transistor [45]. Also, our TCAD simulation results essentially verify this assumption (Fig. 6.6). Based on this assumption, we can generate the equation of surface potential as follow:

$$\varphi_s = V - \frac{qN_d x_d}{\frac{\varepsilon_{si}}{x_d}} = V - \frac{qN_d {x_d}^2}{\varepsilon_{si}} \quad (4)$$

where V represents the electron quasi-Fermi potential at the center of the channel and follows the magnitude of V_{ds} across the channel (See Fig. 6.7). Then, the charge in the channel can be represented by the follow equation:

$$Q = q\varepsilon_{si}x_d = C_{OX}V_{ox} = \varepsilon_{si}E_s \quad (5)$$

where V_{ox} represents the electrical filed across the gate oxide to the surface of the channel and X_d represents the width of depletion region. Since the junction less

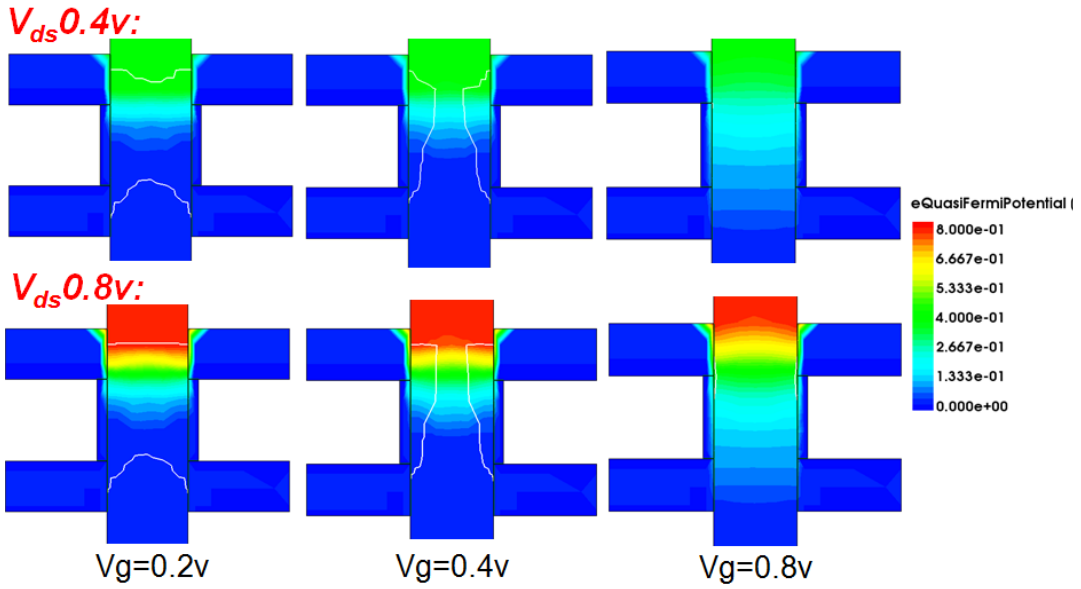

Figure 6.7 **Quasi-Fermi potential distribution at the center of the channel**

transistor operates in the accumulation mode, the relationship between V_{ox} and φ_s follows the equation [46]:

$$V_{ox} = V_G - V_{FB} - \varphi_S \quad (6)$$

where V_G represents the applied gate voltage and V_{FB} represents the flat-band voltage which is the difference of the workfunction of the gate metal and Fermi-level of the doped silicon channel ($\Phi_M - \Phi_S$). The V_{th} is defined at the state where the channel the is just fully depleted. This means the X_d is equal to half of channel width ($W_{ch}/2$). Then we can generate the expression of V_{th}:

$$V_{th} = V_{FB} + \varphi_S + V_{OX} = V_{FB} - \frac{qN_d{x_d}^2}{\varepsilon_{si}} - \frac{qN_d x_d t_{ox}}{\varepsilon_{ox}} = V_{FB} - \frac{qN_d W^2}{4\varepsilon_{si}} - \frac{qN_d W t_{ox}}{2\varepsilon_{ox}}$$

Combined with equation (3)(5)(6), the expression of φ_s is simplified as follow:

$$\varphi_s = V_G - V_{th} - \frac{qN_dW^2}{4\varepsilon_{si}} - v_T W\left[\frac{qN_dW}{4C_{ox}v_T} exp\left(\frac{V_G - V_{th} - V}{v_T}\right)\right] \quad (8)$$

where W is the LambertW-function [47]. Then we can calculate the mobile charge in

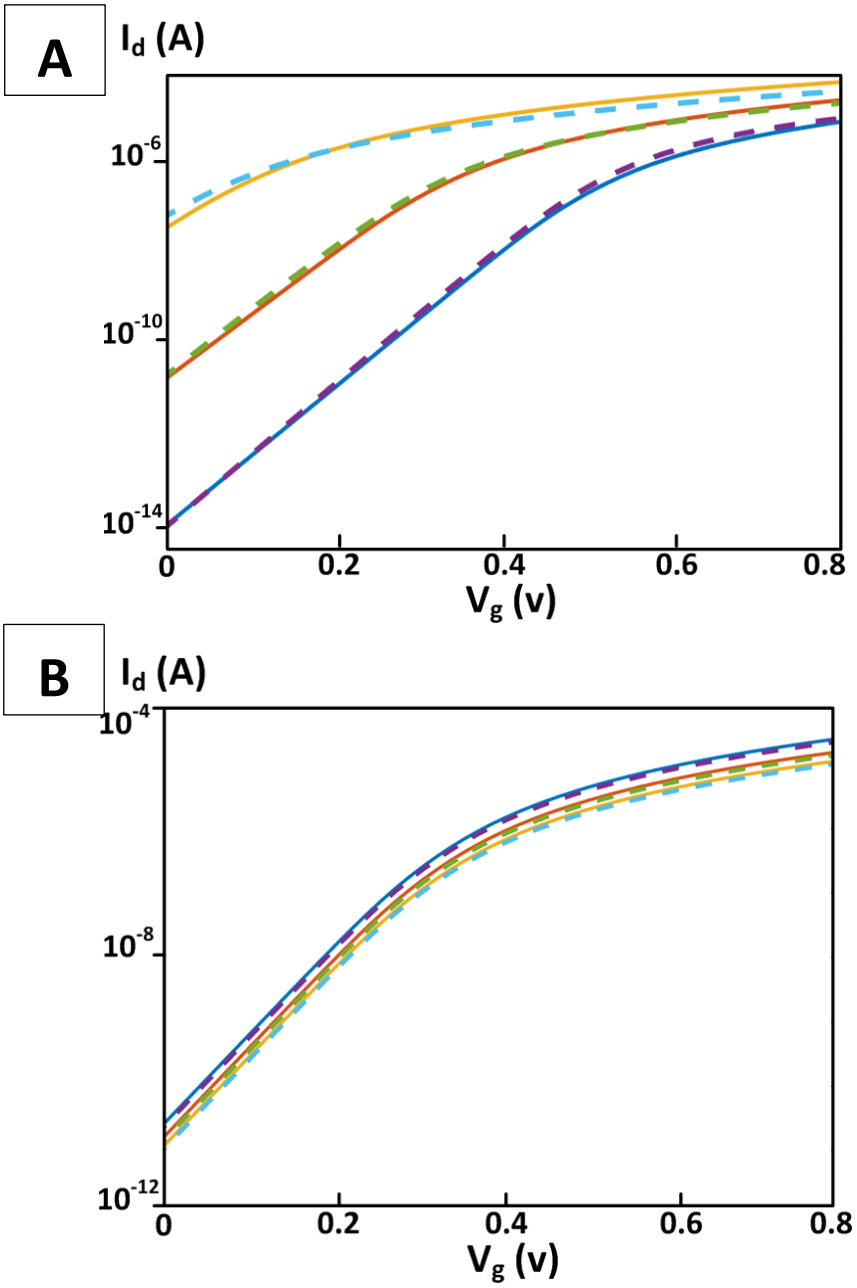

Figure 6.8 **A) Modeled I-V vs. TCAD simulated for various channel widths; B) Modeled I-V vs. TCAD simulated for various channel lengths**

the channel and expressed as:

$$Q_m = qN_d\left(\frac{W}{2} - x_d\right) \quad (9)$$

In this equation, the variable X_d can be generated using equations (4) and (8). Then, the channel current I_{DS} can be expressed as:

$$I_{DS} = \mu \frac{W}{L} \int_0^{V_{DS}} Q_m dV \tag{10}$$

The modeled I-V curve is shown in Fig. 6.8, which shows comparable results with TCAD simulated results. The device capacitance is extracted using TCAD simulation. The value of the parasitic capacitance was assumed to be proportional to the channel length and width [45].

The C-V data of our VGAA junctionless transistors are built based on the TCAD simulations. The C-V data are proportional to the channel size. Firstly, we did capacitance extraction from a VGAA transistor with 16nm width and 16nm channel length. Then, the extracted data was written as look-up table in the Verilog-A model and can be linearly scalable based on the channel length and width.

6.3.2 Device-to-circuit Simulation for SRAM

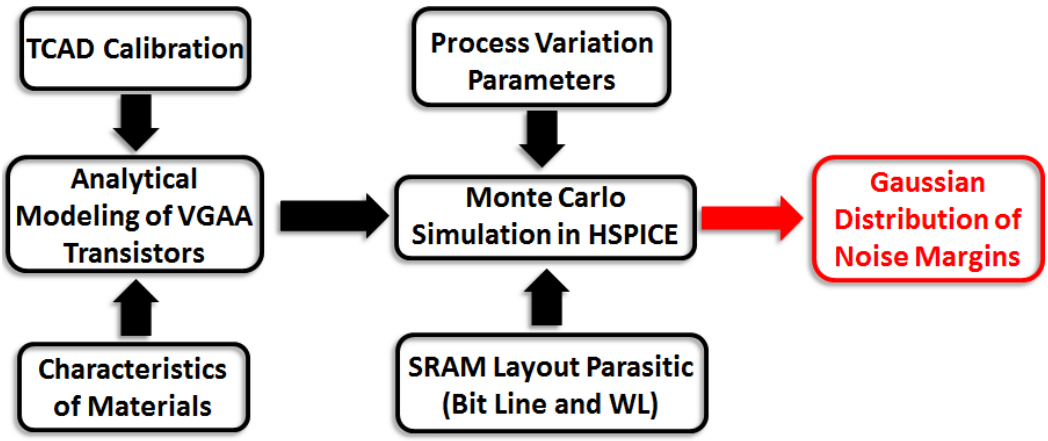

Figure 6.9 **Evaluation flow of variation impact on SRAM**

Fig. 6.9 shows the overall evaluation flow. Monte Carlo simulation was performed

to capture the device's channel length and width variations' impact on SRAM read noise margin. The simulation was performed in HSPICE by importing our analytical VGAA junctionless transistor model, SRAM layout parasitics and geometrical variation parameters of the channel. Details of our device models and SRAM parasitics extraction are shown in Chapter 6.3.1. Since the line width variation comes from lithography variation where the line width data statistically follows a Gaussian distribution, the resulted channel length/width variations and the SRAM noise margin also follow the Gaussian distribution. The mean value μ and stand deviation σ are two key parameters used to express the Gaussian distribution of variation. For the 2D CMOS or M3D, the variations of channel length and width follow the variation distribution in the state-of-the-art lithography technology where the stand deviation σ=8%*μ [49]. For S3DC, the channel width variation is still lithography dependent (σ=8%*μ) but the channel length is deposition dependent (σ=4%*μ[44]). This way, the use of vertical transistor and its unique fabrication enable a significant reduction of channel length variation compared to convention bulk-Si transistor. Also, the characteristics of GAA transistor enable better control on the channel and operation mode which allows the S3DC's SRAM to have better tolerance of variation in channel width. Therefore, despite the S3DC and M3D have the same variation degree in channel width (σ=8%*μ), the VGAA transistor in S3DC would have much less degradation in SS and change in V_{th} compared to the conventional bulk-Si transistor in M3D or 2D CMOS. The detailed evaluation results are shown in chapter 6.3.3.

6.3.3 Variation's Impact on Noise Margin and Failure Rate

Fig. 6.10 shows the Gaussian distribution of read NM for each technology as channel length varies. S3DC's SRAM shows smaller variation compared to M3D as well as 2D CMOS due to the smaller standard deviation σ. Fig. 6.11 shows the

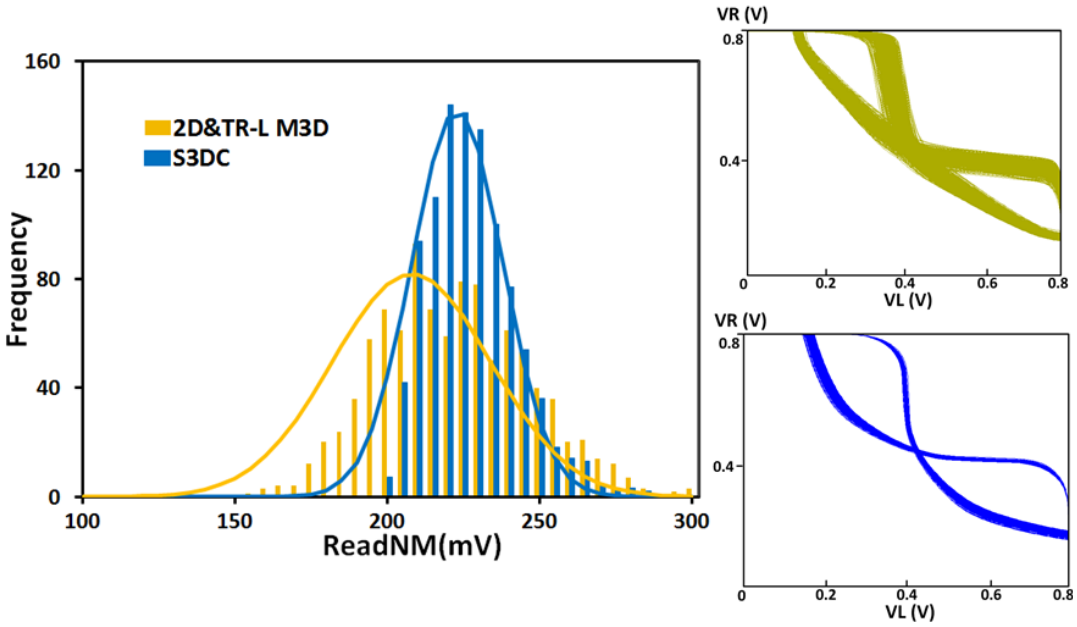

Figure 6.10 **Gaussian distribution of read NM for channel length variation**

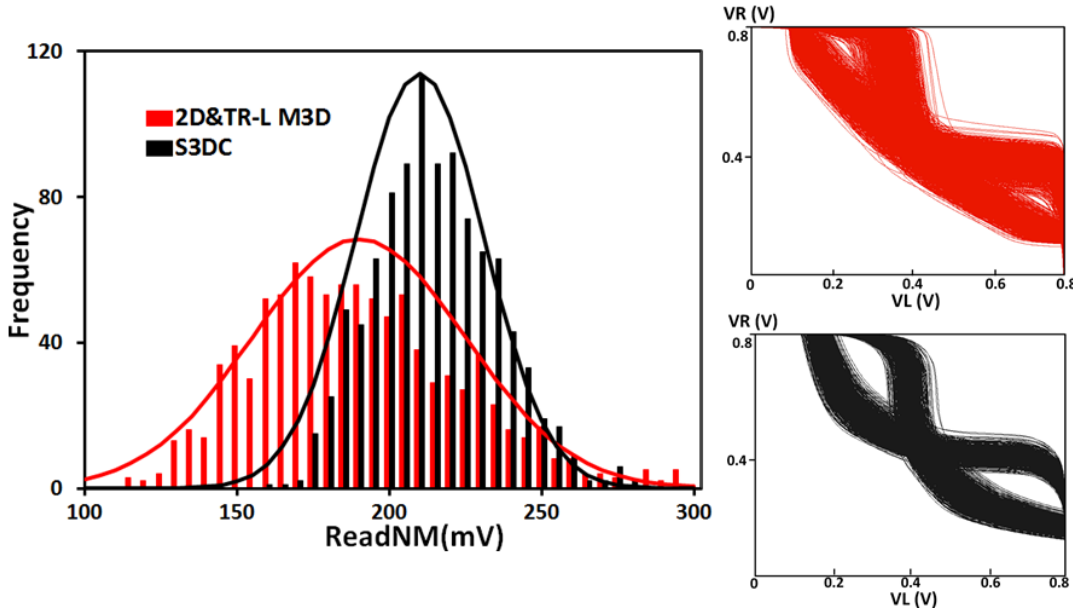

Figure 6.11 **Gaussian distribution of read NM for channel width variation**

Gaussian distribution of read NM for each technology as channel width varies. The

S3DC's SRAM also shows better immunity to variation compared to M3D and 2D CMOS. It should be noted that M3D's SRAM and 2D CMOS based SRAM have the same variation of NM since the NM is device dependent but not design dependent.

The variation in read NM would cause the failure in SRAM. The probability of the

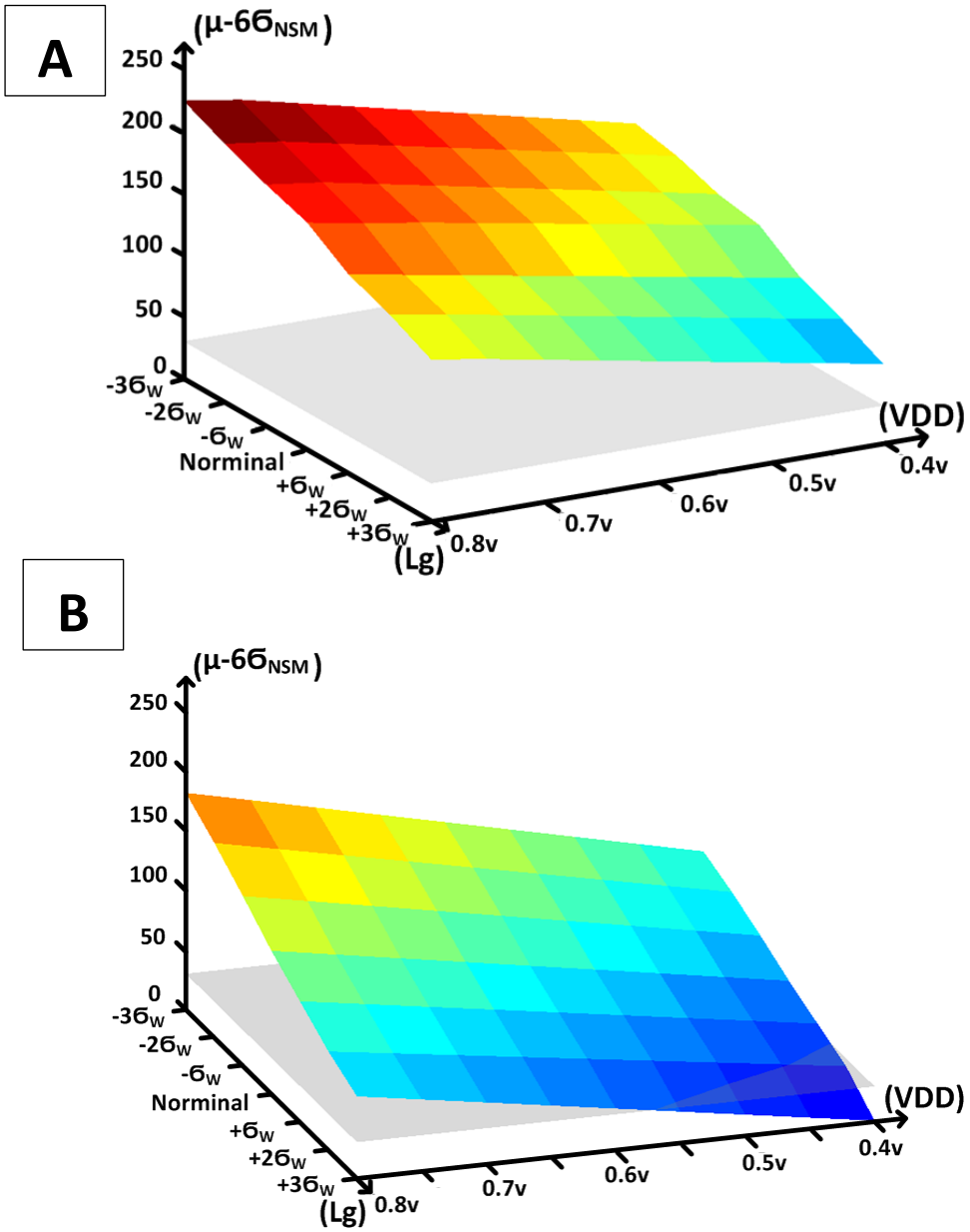

Figure 6.12 **A) (μ-6σ) criterion for channel length variation in S3DC's SRAM as channel width varies from -3σ to +3σ. B) (μ-6σ) criterion for channel length variation in M3D and 2D CMOS based SRAM as channel width varies from -3σ to +3σ.**

failure in SRAM bitcells as the probability at which the read NM is below thermal

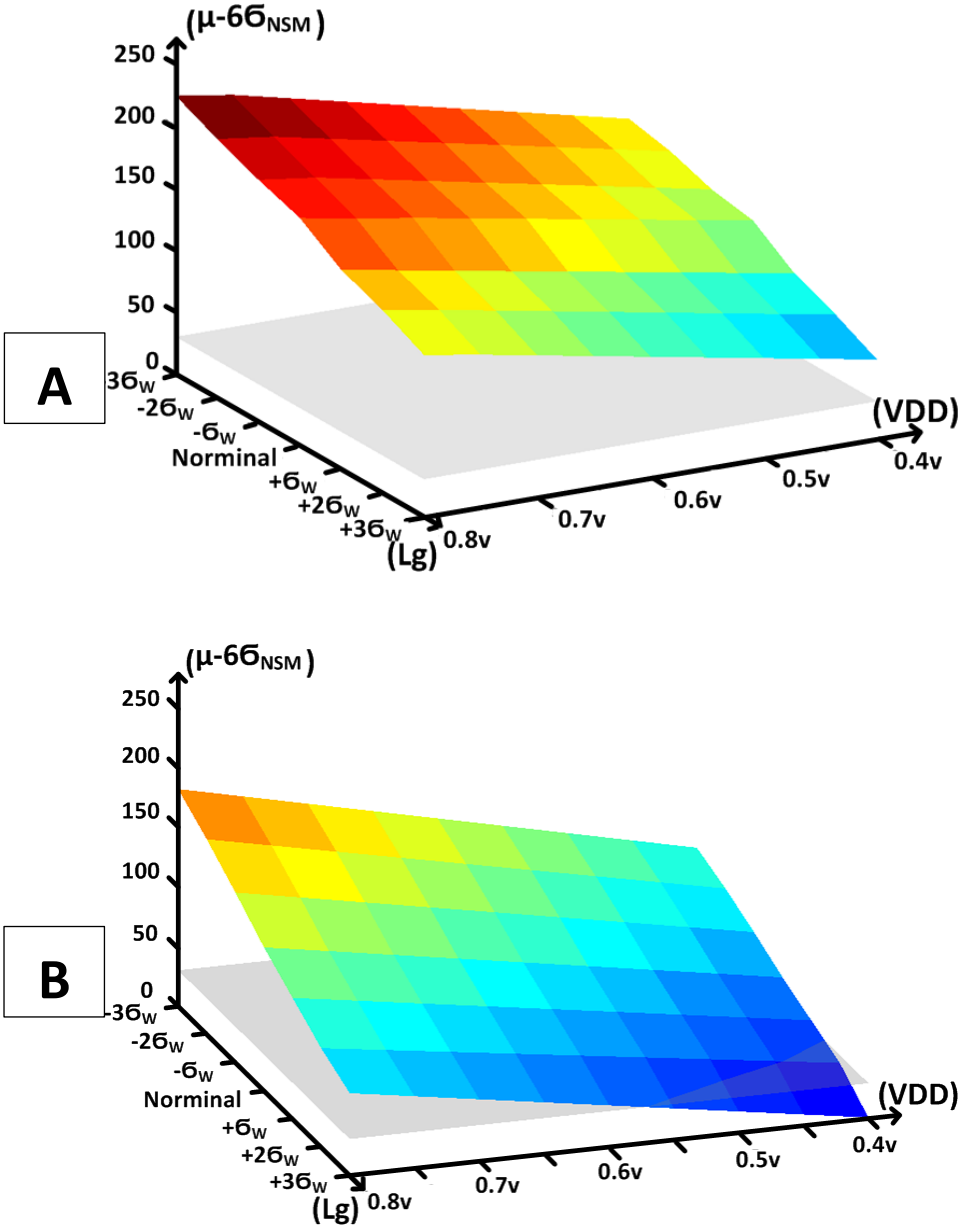

Figure 6.13 **A) (μ-6σ) criterion for channel width variation in S3DC's SRAM as channel length varies from -3σ to +3σ. B) (μ-6σ) criterion for channel width variation in M3D and 2D CMOS based SRAM as channel length varies from -3σ to +3σ.**

noise at 300K [43]. If the read NM is below thermal noise, during the read operation, noise disturbance has a probability to flip the bitcell. To ensure the 6σ yield criterion for the read stability, the worst-case 6σ point away from the mean of the RSNM must lie above the thermal noise (i.e., $\mu_{RSNM} - 6*\sigma_{RSNM} \geqslant 26$ mV). The (μ-6σ) criterion of read NMs in SRAMs in S3DC, TR-L M3D and 2D CMOS for channel length

variation is plotted in Fig. 6.12. Firstly, it can be observed thatS3DC's SRAM is always above 26mV thermal noise while the M3D and 2D CMOS based SRAMs have some parts (low VDD region) that are below thermal noise margin. Secondly, it can be found that as channel width increases from (nominal - 3σ) to (nominal + 3σ), the degradation of NM in S3DC is much less compared to M3D and 2D CMOS. This is caused by the better control of channel in GAA transistor as channel width increases compared to bulk-Si transistor. Fig. 6.13 shows the (μ-6σ) criterion of read NMs in SRAMs in S3DC, TR-L M3D and 2D CMOS for channel width variation. It can be seen that channel width variation has larger impact on NM compared to channel length variation. As shown in 6.13 the SRAMs in both S3DC and M3D have some parts (low VDD region) below thermal noise. But the S3DC's SRAM has overall larger (μ-6σ) criterion than M3D indicating the lower failure rate in S3DC.

BIBLIOGRAPHY

[1] M. Motoyoshi, "Through-Silicon Via (TSV)," in Proceedings of the IEEE, vol. 97, no. 1, pp. 1-4, Jan., 2009.

[2] P. Batude, M. Vinet, A. Pouydebasque, C. Le Royer, B. Previtali, C. Tabone, J.-M. Hartmann, L. Sanchez, L. Baud, V. Carron, A. Toffoli, F. Allain, V. Mazzocchi, D. Lafond, O. Thomas, O. Cueto, N. Bouzaida, D. Fleury, A. Amara, S. Deleonibus, and O. Faynot, "Advances in 3D CMOS sequential integration," in Proceedings of IEEE International Electron Devices Meeting, Washington, 2009, pp. 1-4.

[3] M. S. Ebrahimi, G. Hills, M. M. Sabry, M. M. Shulaker, H. Wei, T. F. Wu, S. Mitra, and H.-S. Philip Wong, "Monolithic 3D integration advances and challenges: From technology to system levels," in Proceedings of SOI-3D-Subthreshold Microelectronics Technology Unified Conference, Millbrae, 2014, pp. 1-2.

[4] Jiajun Shi, Deepak Nayak , Motoi Ichihashi, Srinivasa Banna, and Csaba Andras Moritz, “On the Design of Ultra-High Density 14nm Finfet Based Transistor-Level Monolithic 3D ICs,” in Proceedings of IEEE Computer Society Annual Symposium on VLSI, Pittsburgh, 2016, pp. 449-454.

[5] M. Rahman, S. Khasanvis, J. Shi, M. Li, and C. A. Moritz. (2014, Apr.). Skybridge: 3-D Integrated Circuit Technology Alternative to CMOS. [Online]. Available: http://arxiv.org/abs/1404.0607

[6] M. Rahman, S. Khasanvis, J. Shi, M. Li, and C. A. Moritz, “Fine-Grained 3-D Integrated Circuit Fabric using Vertical Nanowires,” in Proceedings of International 3D System Integration Conference, San Francisco, 2015, pp. TS9.3.1-TS9.3.7.

[7] M. Rahman, S. Khasanvis, J. Shi, M. Li, and C. A. Moritz, "Architecting 3-D Integrated Circuit Fabric with Intrinsic Thermal Management Features," in Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures, Boston, 2015, pp. 157-162.

[8] S. Khasanvis, M. Rahman, M. Li, J. Shi, and C. A. Moritz, “Architecting Connectivity for Fine-grained 3-D Vertically Integrated Circuits,” in Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures, Boston, 2015, pp. 175-180.

[9] M. Rahman, P. Narayanan, S. Khasanvis, J. Nicholson, and C. A. Moritz, “Experimental Prototyping of Beyond-CMOS Nanowire Computing Fabrics,” in Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures, New York, 2013, pp. 134-139.

[10] C.-W. Lee, A. Afzalian, N. D. Akhavan, R. Yan, I. Ferain, and J.-P. Colinge, "Junctionless Multigate Field-Effect Transistor," in Applied Physics Letter, vol. 94, no. 5, pp. 053511, Feb., 2009.

[11] C. Liu and S. K. Lim, "A Design Tradeoff Study with Monolithic 3D Integration," in Proceedings of IEEE International Symposium on Quality Electronic Design, Santa Clara, 2012, pp. 529-536.

[12] S. Panth, et.al., "Design Challenges and Solutions for Ultra- High-Density Monolithic 3D ICs," IEEE S3S, pp. 1-2, 2014.

[13] J. Cong, et.al., "A thermal-driven floorplanning algorithm for 3D ICs," IEEE ICCAD, pp. 306-313, 2004.

[14] J. Shi, et.al., "A 14nm FinFET transistor-level 3D partitioning design to enable high-performance and low-cost monolithic 3D IC," IEDM, pp. 2.5.1 - 2.5.4, 2016.

[15] S. Samal, et.al., "Full Chip Impact Study of Power Delivery Network Designs in Monolithic 3D ICs," IEEE ICCAD, pp. 565 - 572, 2014.

[16] Y. Lee, et.al., "Ultra High Density Logic Designs Using Transistor-Level Monolithic 3D Integration," IEEE ICCAD, pp. 539 - 546, 2012.

[17] S. Bobba, et.al., "CELONCEL: Effective Design Technique for 3-D Monolithic Integration targeting High Performance Integrated Circuits," IEEE ASP-DAC, pp. 336–343, 2011.

[18] M. Li, et.al., "Skybridge-3D-CMOS: A Vertically-Composed Fine-Grained 3D CMOS Integrated Circuit Technology," IEEE ISVLSI, pp. 403-408, 2016.

[19] J. Shi, et.al., "Routability in 3D IC Design: Monolithic 3D vs. Skybridge 3D CMOS" IEEE NANOARCH, pp. 145-150, 2015.

[20] "Synopsys- Sentaurus User Guide," 2015-1.

[21] "Cadence-Voltus IC Power Integrity User Guide, " 2015-8.

[22] "Synopsys- SiliconSmart User Guide," 2015-3.

[23] J. Shi, et.al., "NP-Dynamic Skybridge: A Fine-grained 3D IC Technology with NP-Dynamic Logic", IEEE TETC, Vol. PP, Issue. 99, pp. 1-1, 2017.

[24] J. Shi, et.al., "Architecting NP-Dynamic Skybridge", NANOARCH, pp.169 – 174, 2015.

[25] Nangate. Nangate 15nm Open Cell Library.

[26] K. Choi, et.al. "The Effect of Metal Thickness, Overlayer and High-k Surface Treatment on the Effective Work Function of Metal Electrode “, ESSD ERC, pp. 101-104, 2005.

[27] P. Jiang, et.al. "Dependence of crystal structure and work function of WNx films on the nitrogen content”, Applied Physics Letters, pp. 122107 -122107 -3, 2006.

[28] “LEF/DEF Language Reference” 2011.

[29] “Liberty User Guides and Reference Manual Suite” 2013.

[30] QRC Extraction Users Manual, Cadence, 2010.

[31] Nanoscale Integration and Modeling (NIMO) Group, Arizona State University. (2005). PTM RC Interconnect Models. [Online]. Available: http://ptm.asu.edu.

[32] "Cadence-Encounter User Guide, " 2015-8.

[33] J. Cong, et.al. “Optimizing routability in large-scale mixed-size placement,” ASP-DAC, pp. 441–446, 2013.

[34] P. Saxena, R. S. Shelar, S. S. Sapatnekar, Routing Congestion in VLSI Circuits: Estimation and Optimization, New York: Springer, 2007.

[35] B. S. Landman, et.al. “On a pin versus block relationship for partitions of logic graphs,” IEEE Transactions on Computers, Vol. C-20, Issue. 12, pp. 1469–1479, 1971.

[36] OpenCores. http://opencores.org.

[37] M. Li, et.al., "Skybridge-3D-CMOS: A Vertically-Composed Fine-Grained 3D CMOS Integrated Circuit Technology," IEEE ISVLSI, pp. 403-408, 2016.

[38] J. Shi, et.al., "Routability in 3D IC Design: Monolithic 3D vs. Skybridge 3D CMOS" IEEE NANOARCH, pp. 145-150, 2015.

[39] "Cadence-Voltus User Guide, " 2016-6.

[40] S. Samal, et.al., "Full Chip Impact Study of Power Delivery Network Designs in Monolithic 3D ICs," IEEE ICCAD, pp. 565 - 572, 2014.

[41] A. Kranti, et.al., "Junctionless nanowire transistor: Properties and design guidelines." ESSDERC, pp.357-360, 2010.

[42] C. Liu and S. K. Lim, "Ultra-High Density 3D SRAM Cell Designs for Monolithic 3D Integration," IEEE International Interconnect Technology Conference, pp. 1-3. , 2011.

[43] T. H.-Bao, et.al, "A Comprehensive Benchmark and Optimization of 5-nm Lateral and Vertical GAA 6T-SRAMs", IEEE Transactions on Electron Devices, Vol: 63, Issue: 2, pp. 643 – 651, 2016.

[44] S. S. Sakhare, et.al, "Simplistic Simulation-Based Device-VT-Targeting Technique to Determine Technology High-Density LELE-Gate-Patterned FinFET SRAM in Sub-10 nm Era", IEEE Transactions on Electron Devices, Volume: 62, Issue: 6, pp. 1716 – 1724, 2015.

[45] J. Hur, et.al, " A core compact model for Multiple-Gate Junctionless FETs ", IEEE Transactions on Electron Devices, Vol. 62, Issue. 7, pp. 2285 – 2291, 2015

[46] Chenming Hu," Modern Semiconductor Devices for Integrated Circuits", Upper Saddle River, N.J, 2010.

[47] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function", Adv. Comput. Math., vol. 5, pp. 329–359, 1996.

[48] "Synopsys HSPICE User Guide, " 2013-9.

[49] A. Carlson, et.al, " SRAM Read/Write Margin Enhancements Using FinFETs", Transactions on Very Large Scale Integradation Systems, Vol. 18, Issue. 6, 2010.

[50] H. lwai, "Roadmap for 22 nm and beyond," Microelectronic Eng., vol. 86,pp. 1520–1528, 2009.

[51] M. Rahman, et.al, "Manufacturing pathway and experimental demonstration for nanoscale fine-grained 3-D integrated circuit fabric", IEEE International Conference on Nanotechnology, pp. 1214 – 1217, 2015.

[52] Oracle, "OpenSPARC T2." [Online]. Available: http://www.oracle.com

[53] "Cadence Virtuoso Schematic Composer Tutorial, " Version 6.1.